

Open Classrooms

Rapport d'Analyses Statistiques



FIFA WORLD CUP
Qatar 2022

Jean Lavital

14/05/2022

Sommaire

Remerciements	2
Sources de données	3
Introduction	5
Table des illustrations	6
I. Positionnement des équipes qualifiées	7
A. Evaluation des forces en présence	8
B. Sélection de valeurs pertinentes	11
II. Identification des profils potentiels	18
A. Liens entre les facteurs économiques et sportifs	19
B. Classification des équipes	22
Conclusion	24

Remerciements

Mes remerciements s'adressent en premier lieu à mon Mentor,
Emmanuel Haback II, Data Scientist, pour sa confiance et ses conseils qui m'ont
permis de monter en compétences durant ces 12 derniers mois.

Je tiens également à remercier **Emilie Thoumie**, toute l'équipe pédagogique
d'OpenClassrooms et les intervenants professionnels responsables du parcours
Data Analyst, pour avoir assuré la partie théorique en
partenariat avec l'ENSAE-ENSAI.

Sources de données

Qualifiés pour la Coupe du Monde 2022 de football

Liste des 29 premiers pays qualifiés pour la Phase finale de la Coupe du monde 2022. Avec le nombre d'équipes participants aux qualifications, le nombre de matchs joués, le nombre d'équipes qualifiés ainsi que le nombre de buts inscrits à l'heure actuelle pendant les qualifications 209

: <https://www.fifa.com/fr/tournaments/mens/worldcup/qatar2022/qualifiers>

GDP Per Capita

Produit intérieur Brut exprimé en dollar US \$ en parité de pouvoir d'achat en 2020 pour l'ensemble des pays du monde

: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

Classement FIFA

Classement mondial des équipes nationales de football en fonction du nombre de points FIFA engrangé lors des compétitions internationales, pour l'année 2022

: <https://www.transfermarkt.fr/wettbewerbe/fifa>

Valeur marchande

Dans chacune des équipes, les joueurs sélectionnés ont une valeur estimée sur le marché des transferts. Une valeur marchande qui s'évalue en fonction de critères comme les performances du joueur, son âge, son poste, la durée de son contrat, le championnat dans lequel il évolue ou le montant de ses précédents transferts : Classement mondial des équipes nationales de football en fonction du nombre de points FIFA engrangé lors des compétitions internationales, pour l'année 2022

: <https://www.transfermarkt.fr/wettbewerbe/fifa>

Age moyen par équipe

Classement des équipes par âge moyen, réalisé sur la moyenne de l'âge des 23 joueurs constituant chaque équipe

: <https://www.transfermarkt.fr/brasilien/startseite/verein/3439>

Population mondiale

Nombre d'habitants pour l'ensemble des pays du monde

: <https://donnees.banquemondiale.org/indicateur/SP.POP.TOTL>

Classement FIFA entre 1992 et 2021

Classement mondial des équipes nationales de football en fonction du nombre de points FIFA engrangé lors des compétitions internationales entre 1992 et 2021

: <https://www.kaggle.com/datasets/cashncarry/fifaworldranking>

Nombre de joueurs vainqueurs de la Ligue des Champions (Compétition de Club)

La ligue des champions est la plus prestigieuse des compétitions entre clubs européens

: <https://www.transfermarkt.fr/brasilien/startseite/verein/3439>

Introduction

La Coupe du monde de football 2022 est la 22e édition de la Coupe du monde de football, compétition organisée par la FIFA et qui réunit les meilleures sélections nationales. Elle se déroulera au Qatar du 21 novembre au 18 décembre 2021, jour de la fête nationale du Qatar et une semaine avant Noël, avec une estimation du marché télévisuel potentiel à 3,2 milliards de téléspectateurs. La période inédite à laquelle se déroulera le tournoi est liée au climat du Qatar et aux trop fortes chaleurs y régnant, particulièrement aux dates où se déroule habituellement la compétition (mai-juin-juillet).

C'est une édition de la Coupe du Monde qui est particulière, de par la singularité du pays d'accueil, le Qatar, pays qui n'a jamais participé à une Coupe du Monde de football auparavant et dans le contexte par lequel la compétition se prépare.

Le Qatar manquait d'infrastructures pouvant recevoir ce genre d'évènements, il a fallu donc construire beaucoup et rapidement, en employant un grand nombre d'ouvriers... Des ouvriers sont morts sur les chantiers et d'après certains journalistes et organisations, leur véritable nombre est caché.

La Coupe du monde de 2022 est la septième et dernière édition du Mondial à compter 32 participants (depuis 1998).

À savoir que l'analyse se portera uniquement sur les 29 premiers pays qualifiés et que la Russie suite aux évènements actuels (Intervention armée en Ukraine) a été exclu le 5 Mai dernier par la FIFA et ne participera donc pas au tournoi.

Le choix du Qatar pour la Coupe du Monde 2022 a été fortement critiqué et à l'heure actuelle, on se demande toujours si l'élection n'a pas été corrompue.

Des enquêtes sont toujours en cours dont notamment celle du parquet national financier français qui a ouvert, en 2019, une information judiciaire pour « corruption active et passive », ainsi que pour « recel et blanchiment » de ces délits.

L'étude a pour objectif de proposer une vue d'ensemble afin de pouvoir segmenter toutes les équipes participantes de l'édition 2022 de la Coupe du Monde de Football, en se focalisant sur plusieurs aspects sportifs et économiques

Table des illustrations

[Lien des illustrations Tableau Public](#)

Pays qualifié à travers le Monde / 1.....	9
Evolution du Classement FIFA depuis 2010 / 2	11
Age moyen par équipe / 3	12
Nombre de joueurs évoluant à l'étranger et ayant déjà gagné une fois la	
Champions League / 4	13
Population en 2020 / 5	14
Produit intérieur brut / habitant / 6	15
Valeur marchande par équipe / 7	16
Cercles des corrélations / 8	19
Projection des individus / 9	20
Dendrogramme / 10	22
Heat-Map / 11.....	23

I. Positionnement des équipes qualifiées

A. Evaluation des forces en présence

1) Description du déroulement de la phase de poule de la compétition

Le coup d'envoi de la prochaine édition de la Coupe du Monde de Football approche.

Le 21 Novembre, débutera la compétition avec les phases de groupes qui opposeront 8 groupes de 4 équipes.

29 équipes sont d'ores et déjà qualifiées pour cette phase finale.



The infographic displays the 16 teams of the 2022 FIFA World Cup, organized into eight groups (A through H). Each team is represented by its national flag and name. The background is a dark red with a subtle pattern of soccer balls. The FIFA World Cup Qatar 2022 logo is in the top left corner.

Group A	Group B	Group C	Group D
QATAR	ENGLAND	ARGENTINA	FRANCE
ECUADOR	IRAN	SAUDI ARABIA	PER/AUS/UAE
SENEGAL	USA	MEXICO	DENMARK
NETHERLANDS	UKR/SCO/WAL	POLAND	TUNISIA

Group E	Group F	Group G	Group H
SPAIN	BELGIUM	BRAZIL	PORTUGAL
CRC/NZL	CANADA	SERBIA	GHANA
GERMANY	MOROCCO	SWITZERLAND	URUGUAY
JAPAN	CROATIA	CAMEROON	SOUTH KOREA

Il reste seulement 3 places qualificatives pour la phase finale.

- Dans le Groupe B, l'Ukraine, l'Écosse ou le Pays de Galles obtiendra un billet pour la phase finale
- Dans le Groupe D, la dernière place se jouera entre le Pérou, l'Australie et les Émirats arabes-unis
- Dans le Groupe E, une place reste à prendre pour le Costa Rica ou la Nouvelle Zélande

La précédente édition qui s'est déroulée en Russie, a vu la France, une des équipes favorites, s'imposer face à la Croatie. 2018 aura été l'année de la France pour l'édition passée.

a) Coup d'œil sur les équipes présentes

On retrouve parmi les 29 premiers qualifiés pour cette 7^{ème} édition de la Coupe du Monde de Football, les 3 derniers vainqueurs à savoir L'Espagne en 2010, l'Allemagne en 2014 et la France en 2018.

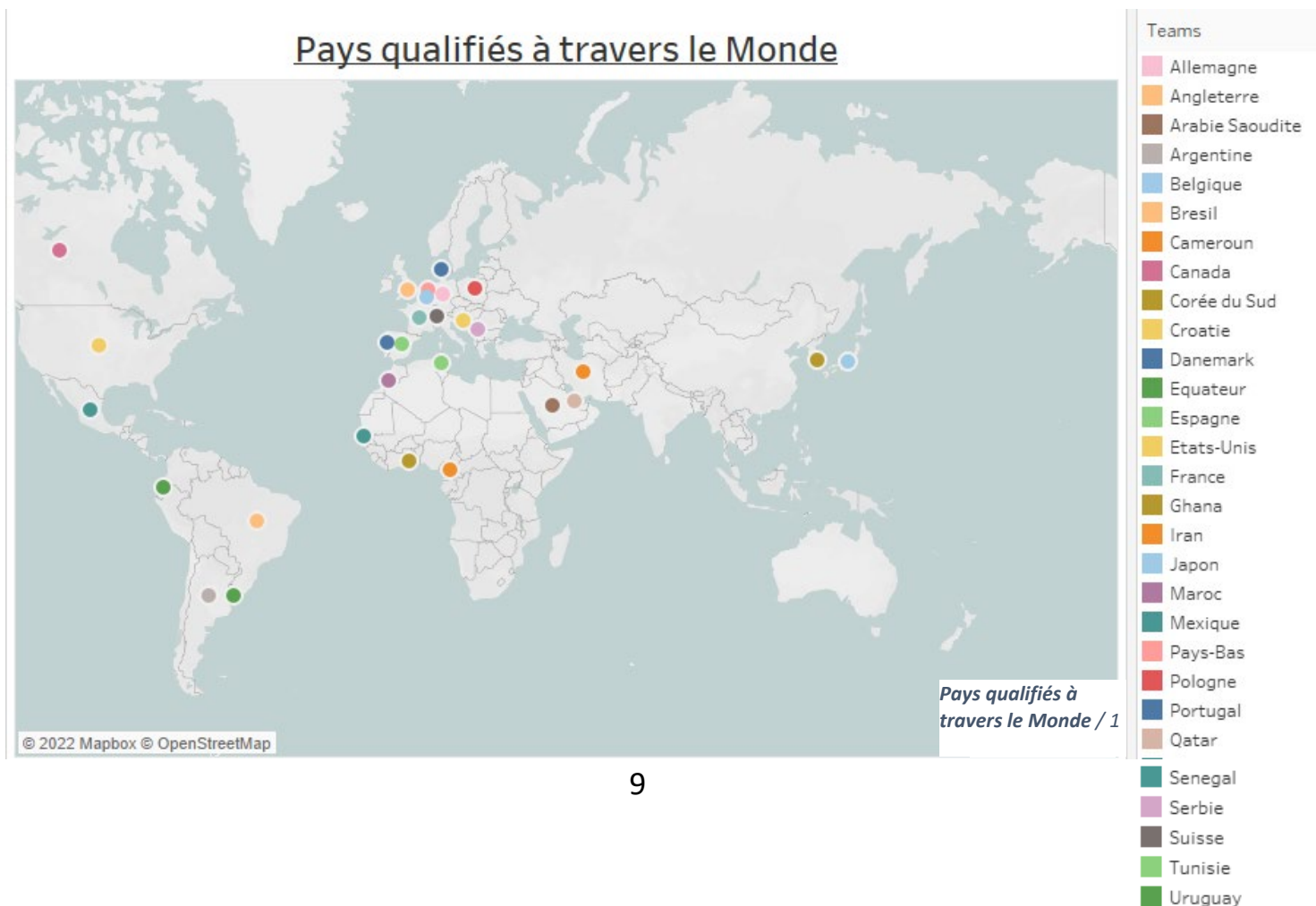
Toutes les principales grandes nations du Football sont présentes pour ce rendez-vous (Le Brésil, l'Argentine, Le Portugal etc...) mis à part une surprise, l'Italie, récente vainqueur de l'Euro 2020, qui s'est faite éliminer en barrages face à la Macédoine du Nord.

On peut également noter la présence du Sénégal, récent champion d'Afrique

Le Qatar est donc le pays hôte de cette Coupe du Monde 2022 de Football, pays pour lequel c'est la première participation.

Le Football est une vitrine pour le Qatar, qui aura à cœur pendant la compétition de montrer une image rayonnante de son savoir-faire et pourquoi pas briller au sein de la compétition avec son équipe nationale.

Ci-dessous, une carte représentant géographiquement, les 29 premiers pays qualifiés pour la phase finale.



- L'Europe est la zone du monde la plus représentée avec 10 pays (France, Allemagne, Angleterre, Suisse, Danemark, Espagne, Portugal, Serbie, Pologne, Pays-Bas).
- L'Afrique aura quant à elle, 5 représentants (Sénégal, Tunisie, Maroc, Ghana, Cameroun).
- L'Amérique du Nord, 3 avec le Canada, Les Etats-Unis et le Mexique
- L'Amérique du Sud, 4 avec Le Brésil, L'Argentine, L'Equateur, L'Uruguay
- L'Asie avec deux pays (Japon et Corée du Sud)
- L'Orient, 3 avec L'Iran, le Qatar et l'Arabie Saoudite

Ce format de compétition composé de 32 équipes pour la phase finale sera le dernier du genre car lors d'édition 2026, 48 équipes se retrouveront en phase finale de la Coupe du monde, à la suite de la décision de la FIFA, promue et adoptée le 10 janvier 2017, d'élargir le tournoi de 32 à 48 participants. Au premier tour les 48 participants seront répartis dans 16 groupes de trois équipes, dont deux se qualifieront pour les 16^e de finale à partir desquels tous les matchs seront à élimination directe.



B. Sélection de valeurs pertinentes

De manière à réaliser un ciblage et grouper les équipes en fonctions de leurs ressemblances, la sélection de différentes variables avec des aspects sportifs et économiques va être déterminante.

1) Des Variables sur l'aspect sportif

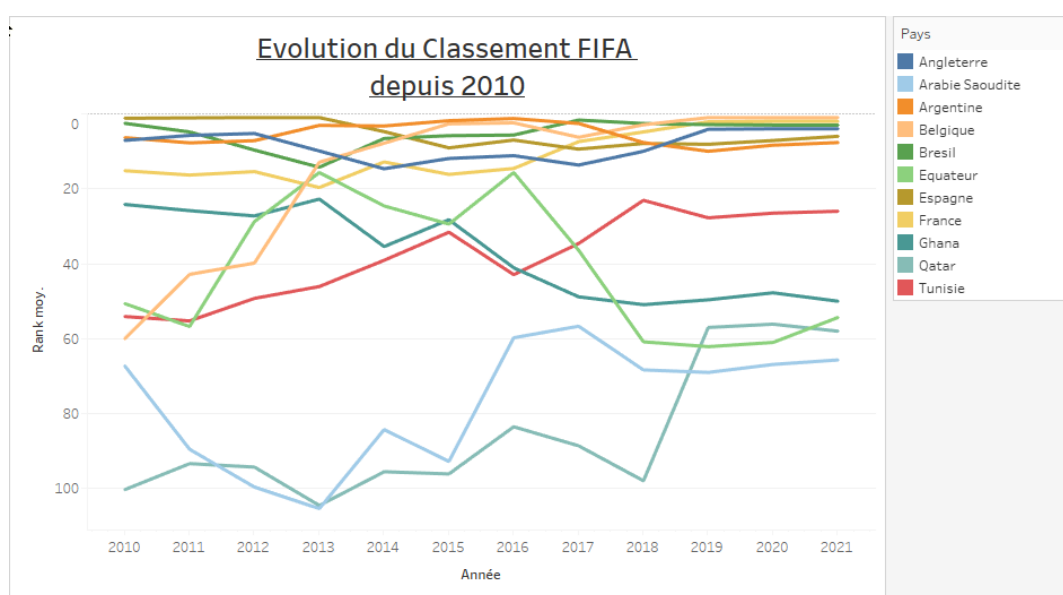
a) Le Classement FIFA

Le Classement FIFA est l'indice de référence pour comparer les meilleures équipes du monde entier et ainsi déterminer la place de celles-ci.

Il permet selon un aspect purement sportif de déterminer selon les résultats des équipes sur les dernières compétitions internationales de leur attribuer des points et donc de leur attribuer un classement par rapport aux autres équipes.

Depuis Août 2018, la méthode de calcul du Classement FIFA a évolué.

Le précédent Classement mondial se basait sur une moyenne, établie sur une certaine période de temps. Le nouveau système procède par addition/soustraction sur un total de points, en fonction des matches gagnés ou perdus. Les points ajoutés ou soustraits sont en partie définis par la force relative des deux adversaires. Le modèle prend donc en compte la logique qui voudrait qu'une équipe mieux classée obtienne de meilleurs résultats face à un adversaire qui occupe une position inférieure au classement.



Evolution du Classement FIFA depuis 2010 / 2

On note la présence régulière aux premières places du classement de nations comme la France, l'Angleterre, L'Argentine, Le Brésil, L'Espagne depuis 2010. La meilleure progression depuis 2010 est l'œuvre de la Belgique qui est passé de la 60^{ème} à la 1^{ère} place du classement en l'espace d'une dizaine d'année. Une progression régulière associée à de bons résultats lors des dernières épreuves internationales et cela sans victoire majeure lors d'une compétition.

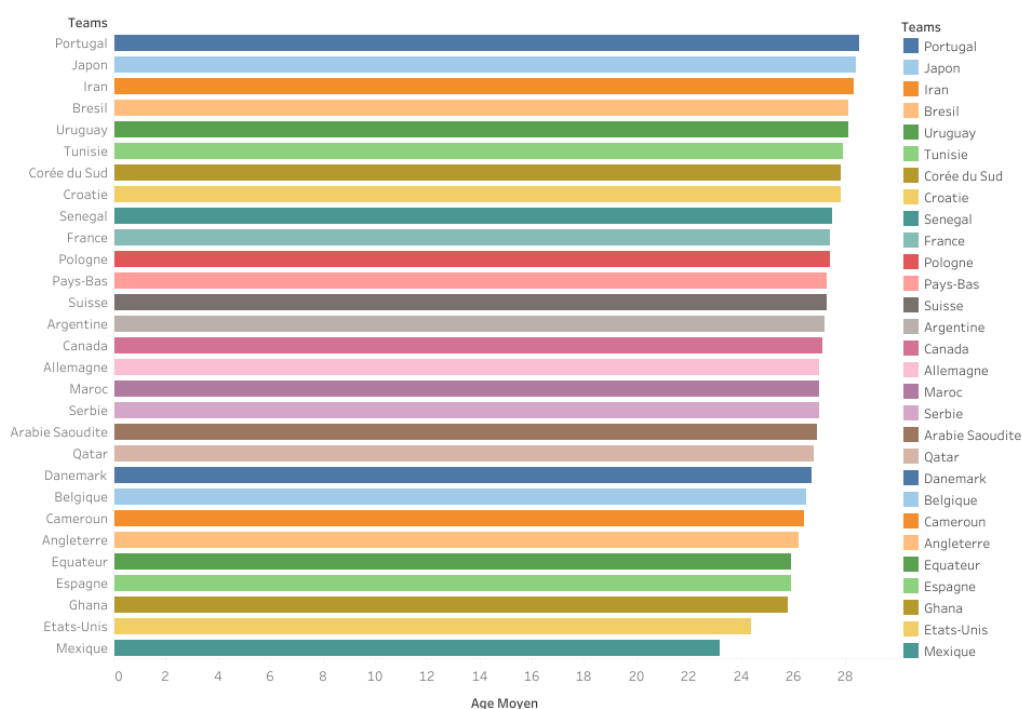
La France est le champion du Monde en titre. L'Espagne a remporté l'Euro 2008, La Coupe du Monde 2010 et encore un Euro en 2012.

La Tunisie connaît également une progression notable sur une dizaine d'année en gagnant près de 20 places au classement (de la 45^{ème} à la 25^{ème})

b) Age Moyen par équipe

La moyenne d'âge peut donner une indication sur l'expérience engrangée par une équipe de football, c'est également un facteur qui peut faire la différence face à d'autres équipes avec une moyenne d'âge plus faible ou plus élevée. Elle est révélatrice de la volonté d'un entraîneur de se lancer avec un groupe confirmé ou rajeuni pour entamer un nouveau cycle de plusieurs années.

AGE MOYEN PAR EQUIPE



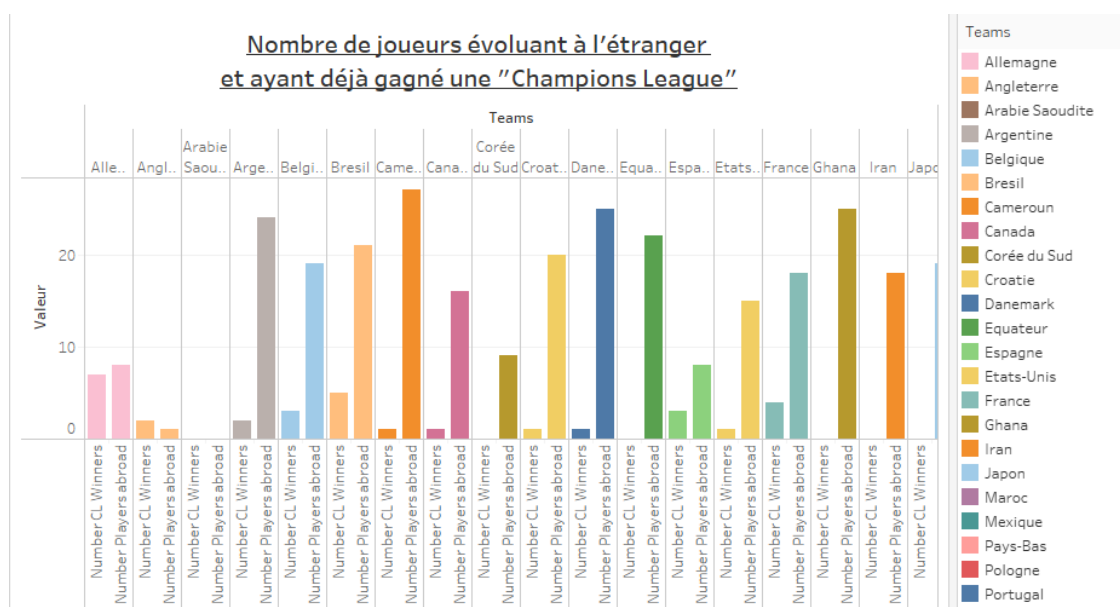
Le Portugal est l'équipe avec la moyenne d'âge la plus élevée (près de 28 ans), alors que le Mexique détient la moyenne d'âge la plus basse avec près de 23 ans. On retrouve les Etats-Unis avec une moyenne d'âge parmi les plus basses avec un âge moyen de 24,6 ans.

Une nouvelle génération de joueurs émerge en sélection avec le travail produit par la fédération de football américaine depuis une quinzaine d'année. De plus en plus de joueurs viennent désormais jouer dans des clubs européens et viennent ainsi apprendre en s'acculturant du football occidental. On peut signaler aussi que sur les équipes victorieuses des éditions précédentes, on retrouve des moyennes d'âge allant de 26 à 29 ans pour l'effectif.

c) Nombre de Joueurs évoluant dans des clubs étrangers par équipe / Nombre de Champions League gagné au moins une fois par équipe

Le nombre de joueurs composant chaque sélection présente parmi les participants à la prochaine Coupe du Monde s'élève à 23. On peut voir à travers ces 23 joueurs, le nombre de joueurs qui jouent dans une ligue autre que la ligue nationale du pays qu'ils représentent. C'est un nombre qui peut varier en fonction du niveau et de la puissance financière de leur ligue nationale.

Le Nombre de Champions League gagné au moins une fois par un joueur dans l'équipe nationale qu'il représente peut s'avérer significatif du point de vue sportif et peut relater d'un certain niveau de jeu et d'expérience de l'effectif si son nombre est plus ou moins élevé



Nombre de joueurs évoluant à l'étranger et ayant déjà gagné une fois la Champions League / 4

En prenant l'exemple d'un pays comme L'Allemagne, la nation allemande détient quelques éléments dans son groupe avec un gros palmarès individuel et fait partie de l'un des 5 plus grands championnats nationaux au monde.

L'équipe nationale allemande de football n'a que 8 joueurs qui évoluent en dehors du championnat allemand (Bundesliga) et compte dans ses rangs 7 joueurs qui ont déjà remporté au moins une fois la plus prestigieuse des compétitions internationales de club : La Ligue des Champions (Champions League).

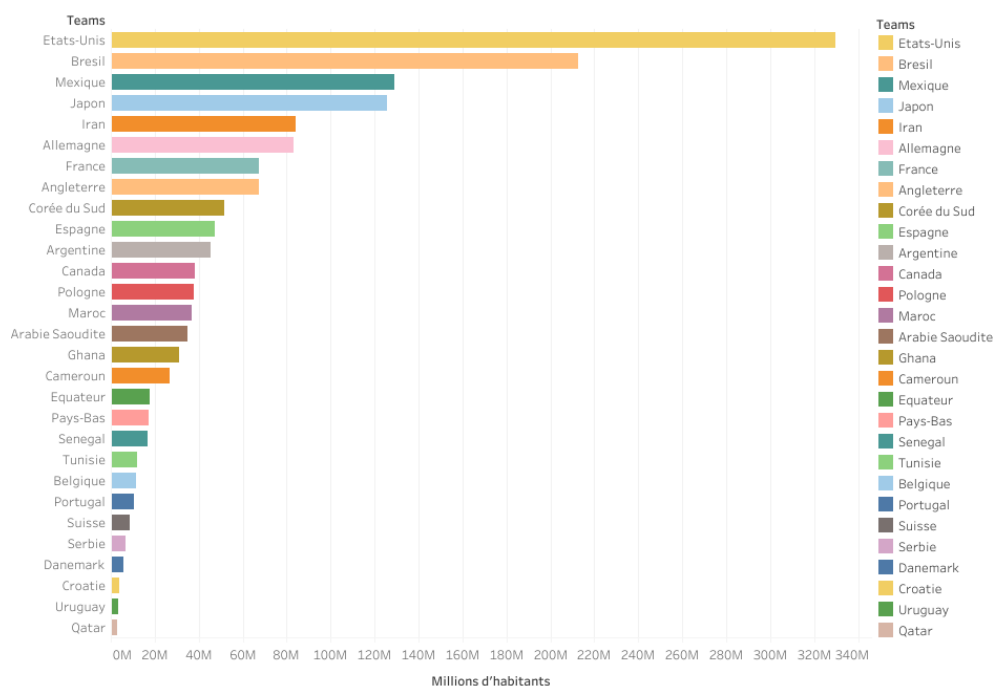
Le Cameroun quant à lui détient seulement un joueur qui a remporté la Ligue des Champions mais a l'intégralité de son groupe qui évolue dans un championnat à l'étranger (27)

2) Des variables sur l'aspect économique

a) Population

La population d'un pays peut avoir une influence sportive et à la fois économique. Plus un pays détient d'habitant et plus le choix sportif sera grand quant à la sélection des joueurs pour constituer un groupe capable de se mesurer aux autres nations lors de compétitions internationales. Il y a une influence sur le niveau des joueurs, ce qu'on appelle dans le jargon « le vivier »

POPULATION EN 2020



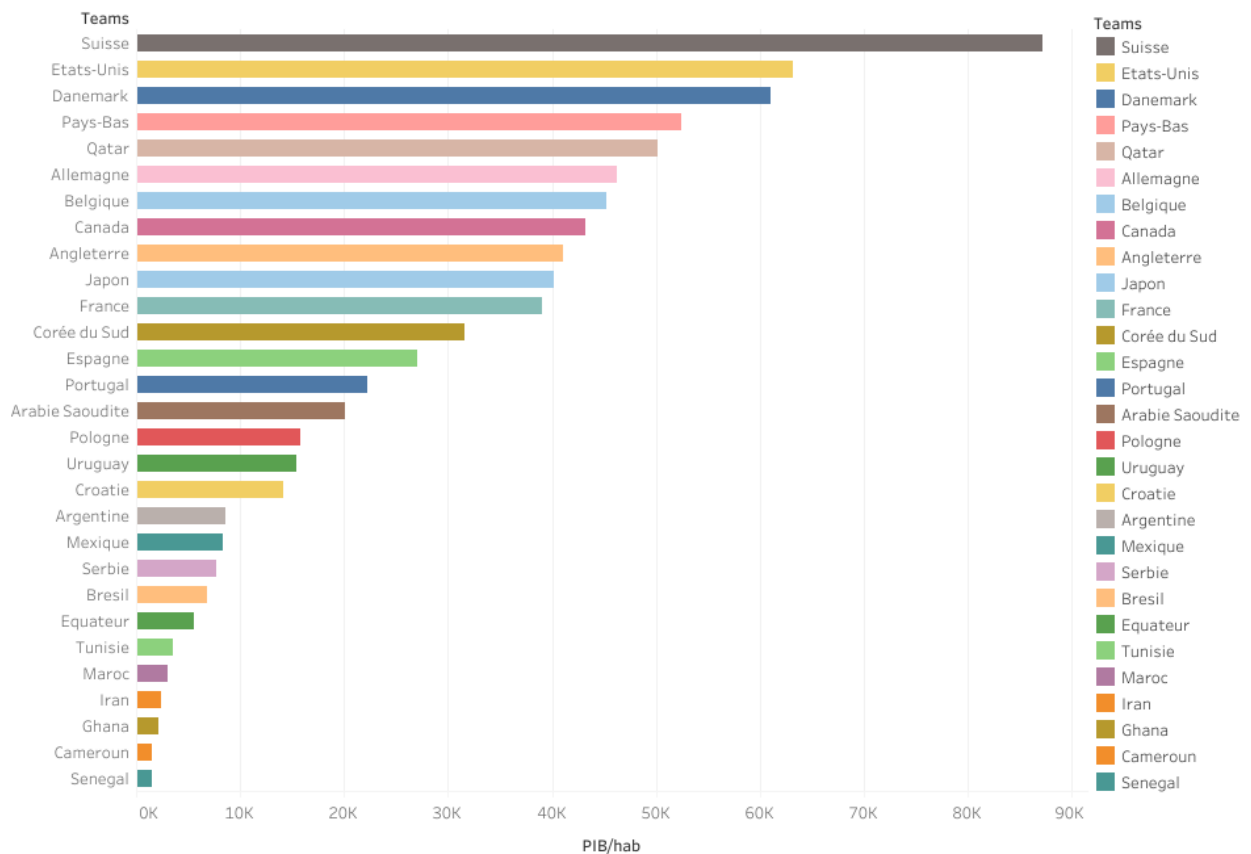
Les pays détenant les plus grandes populations sont principalement des pays issus des continents Américains. Les Etats-Unis avec près de 330 millions d'habitants qui est considéré comme une nation en devenir du football mondial, Le Brésil qui est avec 210 millions d'habitants, une très forte place du football mondial quant à son vivier de joueurs et le Mexique avec près de 130 millions d'habitants.

On retrouve les pays européens entre 5 et 80 millions d'habitants, les pays africains entre 20 et 40 millions d'habitants, les pays d'Asie comme le Japon et la Corée du sud avec respectivement près de 125 et 65 millions et le Qatar en dernière position avec près de 3 millions d'habitants.

b) Produit intérieur brut / habitant

Le produit intérieur brut représente la puissance économique d'un pays, il est significatif pour effectuer une comparaison avec d'autres pays

PRODUIT INTERIEUR BRUT / HABITANT



Produit intérieur brut / habitant / 6

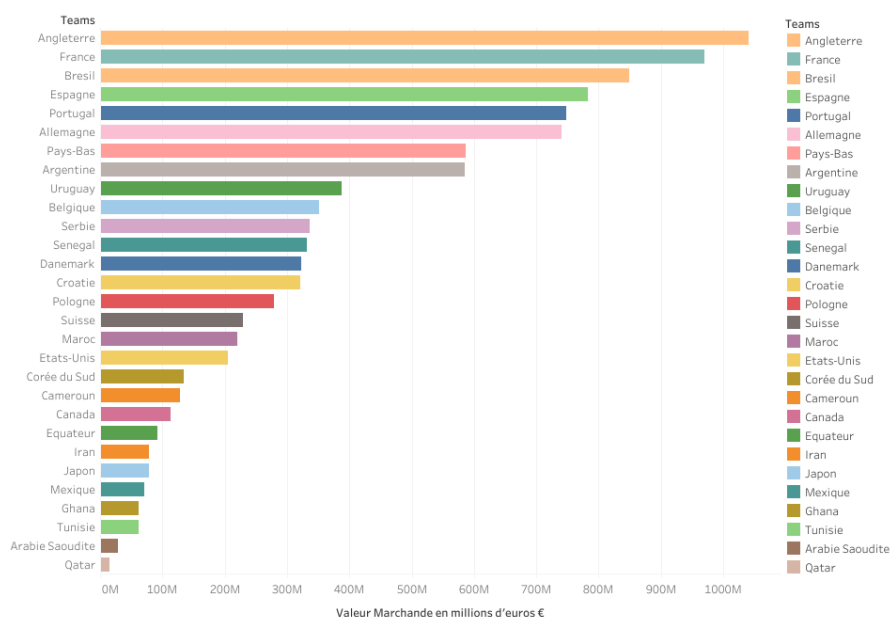
La Suisse colle en tête de ce classement avec plus de 80 000€ / habitant. On retrouve la majorité des pays européens aux premières positions avec les pays d'Amérique du Nord. Le Qatar avec sa puissance financière est également aux avant-postes. Les pays africains et d'Amérique du sud occupent quant à eux, les dernières positions de ce classement.

c) Valeur marchande par joueur cumulée

Plusieurs facteurs sont à prendre en compte pour calculer la valeur marchande d'un joueur de football.

- 1) Son âge : plus il est jeune, plus le potentiel de progression d'un joueur est important. A 23 ans, tout le contraire d'un joueur âgé de plus de 30 ans, qui offre peu de perspectives de plus-value.
- 2) La durée restante de son contrat et les revenus sponsoring : plus la durée du contrat s'amenuise (moins de deux ans), plus le pouvoir de négociation du club propriétaire est faible.
- 3) Le poste du joueur : les attaquants et joueurs réputés décisifs sont plus prisés que les défenseurs et les gardiens de but.
- 4) Le statut en club et en sélection : Sont pris en compte le club et le championnat où se produit le joueur mais aussi son statut (titulaire ou remplaçant) ainsi que ses performances, comme le nombre de buts inscrits.

VALEUR MARCHANDE PAR EQUIPE



Valeur marchande par équipe / 7

La puissance financière du championnat anglais (Premier League) grâce à ses revenus télévisuels et marketing en font le championnat n°1 dans le monde à la lutte avec le championnat espagnol (Liga). On retrouve en position n°1, L'Angleterre qui détient des joueurs à très grosse valeur marchande et qui sont également grassement payés par les clubs anglais.

En seconde position, on retrouve la France qui s'appuie sur une qualité de formation remarquable dont des clubs de toute l'Europe entière viennent recruter sur le marché français. Le joueur français sait s'adapter, n'est pas cher et peut souvent avoir une grosse marge de progression. Les derniers résultats de la sélection française jouent un rôle dans la valeur marchande des joueurs français. Pratiquement 90% de la sélection joue dans des clubs de très haut rang.

Pour les pays en bas de tableau, ils ne détiennent tout simplement pas encore assez de joueur qui puissent leur permettre de redresser le tir mais cela ne veut pas forcément dire que leurs joueurs ne sont pas de qualité ou n'ont aucun potentiel.

Avec cette évaluation des forces en présence de par les différentes variables utilisées pour décrire les équipes qualifiées pour la phase finale de la compétition, on peut désormais tenter de les rassembler et de déterminer des profils et groupes entre elles.

II. Identification des profils potentiels

A. Liens entre les facteurs économiques et sportifs

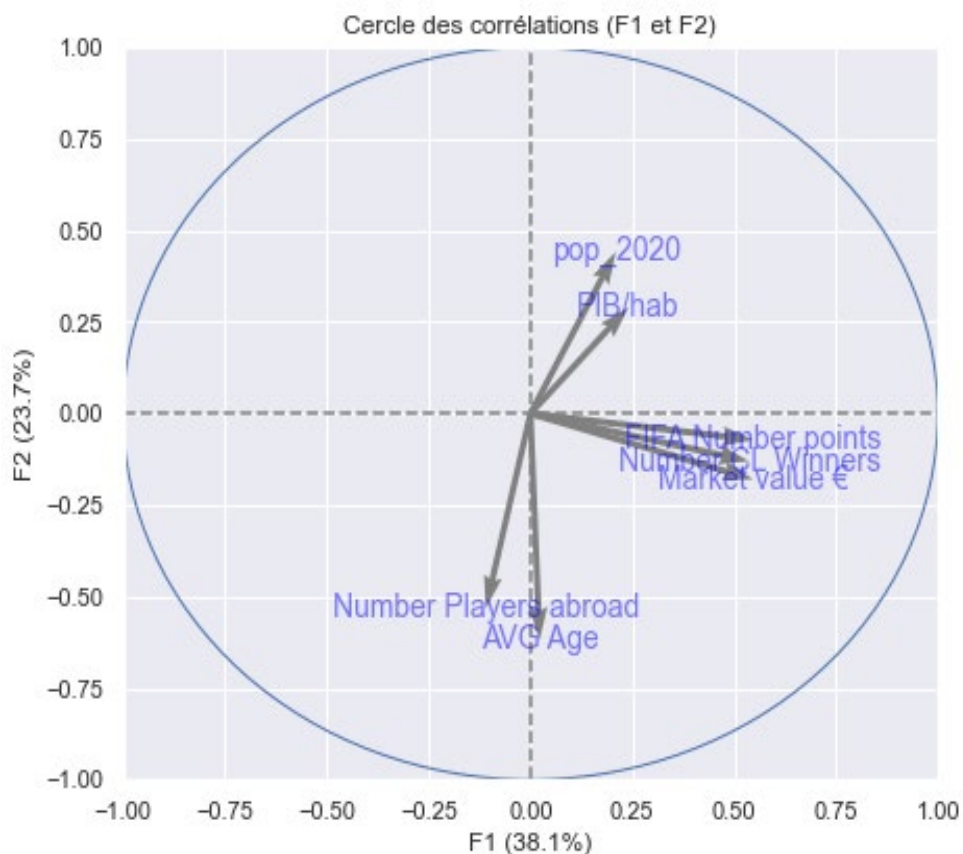
1) Analyse en composantes principales (ACP)

Une Analyse en composantes permet de comprendre la structure d'un ensemble de variables, créer des instruments pour analyser des éléments impossibles à mesurer directement, et condenser les informations issues d'un grand nombre de variables dans un ensemble restreint en garantissant une perte minimale.

Elle permet également :

- D'identifier des profils cachés dans un jeu de données
- De réduire les dimensions des données en enlevant la redondance des données
- D'identifier des variables corrélées

En utilisant le module « *Scikit-Learn* » sur Python, on parvient à ce résultat pour la réalisation du premier cercle de corrélation :



Cercles des corrélations / 8

D'après le cercle de corrélation des composantes F1 et F2 :

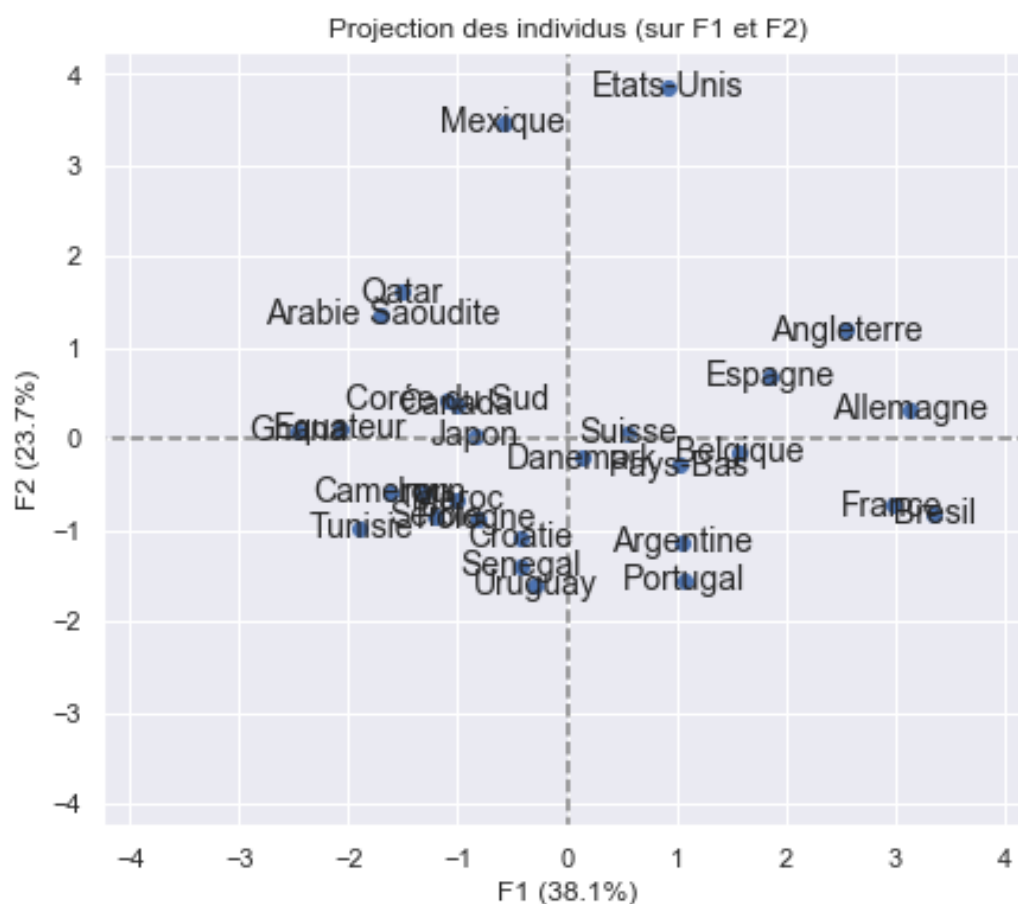
Les variables « *Market Value* », « *FIFA Number points* » et « *Number CL Winners* » sont fortement corrélées. Ces 3 variables ont un coefficient de corrélation de près de 0,5 par rapport à l'axe des abscisses F1

Les variables « *pop_2020* » et « *PIB/hab* » sont corrélées positivement à l'axe des abscisses F1 avec respectivement des coefficients de corrélations de 0,2 et 0,25. Elles sont également corrélées positivement à l'axe des ordonnées F2 avec des coefficients de corrélations de 0,4 et 0,25.

Les variables « *Number Player abroad* » et « *AVG Age* » sont corrélées négativement à l'axe des ordonnées F2 avec des coefficients de -0,5 et -0,55

Les variables variables fortement corrélées que sont « *Market Value* », « *FIFA Number points* » et « *Number CL Winners* » s'associent au composant F1 qui pourrait être qualifié de « Favoris » pour la Coupe du Monde.

2) Projections des individus



Projection des individus / 9

Selon la projection des individus obtenus avec les axes F1 et F2, le Cercle des corrélations est bien respecté avec le positionnement des équipes dites « Favoris » selon la composante principale F1 rebaptisée.

L'Angleterre, La France, Le Brésil, l'Allemagne, la Belgique, L'Argentine, Le Portugal, La Suisse, Le Danemark et les Pays-Bas sont toutes positionnées sur le côté droit de l'axe F1, des équipes qui sont en adéquations avec les 3 variables corrélées précédemment identifiées.

Le positionnement des Etats-Unis et du Mexique est expliqué par leur forte population, le Mexique se situe sur la partie gauche de l'axe F1 car le nombre de joueur évoluant à l'étranger pour la sélection mexicaine est faible : seulement 3 joueurs.

Les équipes positionnées sur la partie inférieur gauche du graphique sont des équipes dont la majorité des joueurs évoluent à l'étranger et dont les populations respectives sont plus petites.

Exemple avec la Tunisie qui compte dans ses rangs 21 joueurs qui jouent dans des championnats étrangers et « seulement » près de 12 millions d'habitants.

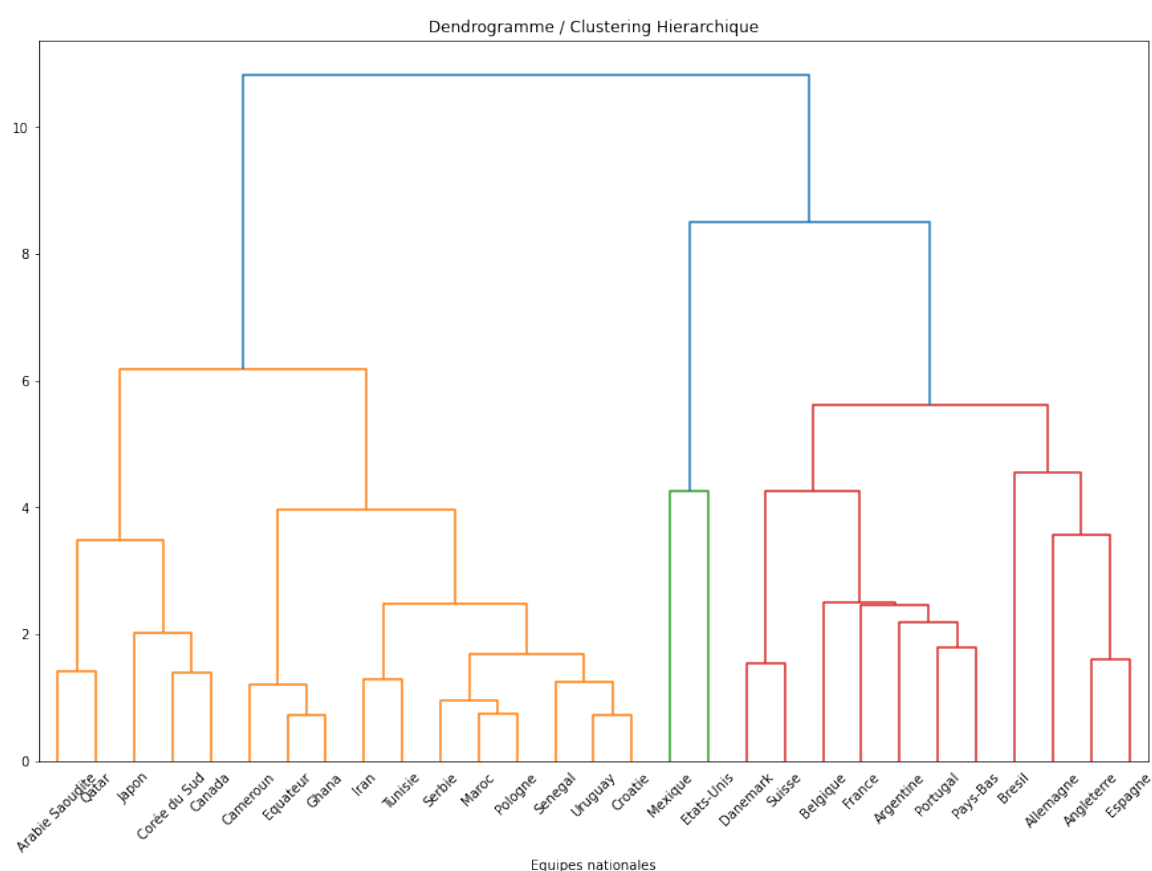
B. Classification des équipes

1) Clustering Hiérarchique

a) Dendrogramme

Le Dendrogramme est un type de diagramme en arborescence que l'on utilise pour présenter le clustering hiérarchique, à savoir les relations entre des ensembles de données similaires.

Le clustering hiérarchique va permettre de différencier les équipes participantes en groupes.



Dendrogramme / 10

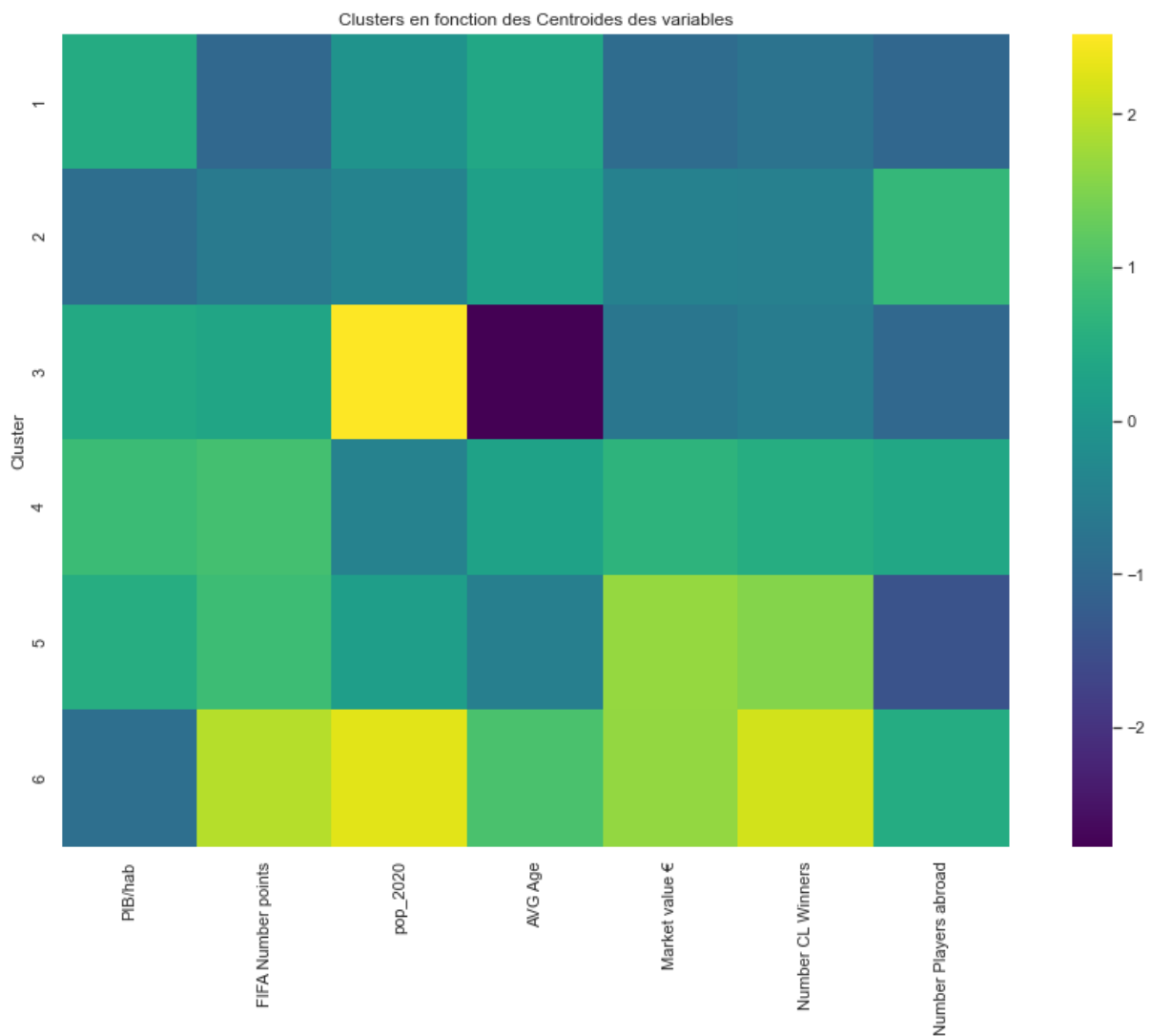
Le Dendrogramme a classé les équipes en 6 clusters :

- **Le Cluster 1** : Arabie Saoudite, Qatar, Japon, Corée du Sud, Canada
- **Le Cluster 2** : Cameroun, Equateur, Ghana, Iran, Tunisie, Serbie, Maroc, Pologne, Sénégal, Uruguay, Croatie
- **Le Cluster 3** : Mexique, Etats-Unis
- **Le Cluster 4** : Danemark, Suisse, Belgique, France, Argentine, Portugal
- **Le Cluster 5** : Allemagne, Angleterre, Espagne
- **Le Cluster 6** : Brésil

b) K-Means et Heat-Map

Le K-Means est l'un des algorithmes de clustering les plus répandus. Il permet d'analyser un jeu de données caractérisées par un ensemble de descripteurs, afin de regrouper les données "similaires" en groupes (ou clusters).

A l'aide du positionnement des centroïdes des variables, ci-dessous on peut voir le résultat d'une « Heat-Map »



Heat-Map / 11

Les Clusters 4, 5 et 6 ont des niveaux de corrélations dans les variables à aspect sportif supérieur à 1.

Le Cluster 6 : Le Brésil est bien corrélé avec la majorité des variables.

Conclusion

Le Rapport d'analyses présent sur les 29 premières équipes qualifiés pour la phase finale de la Coupe du Monde de Football 2022 a été réalisé avec certaines données et variables. Des données et variables supplémentaires auraient pu être additionnés à celles déjà présentes aux vues de l'immensité de ces dernières.

Les Clusters produits par le K-Means et le Dendrogramme ont révélés 6 groupes d'équipes allant des plus faibles aux plus fortes.

Les équipes des clusters 4, 5 et 6 grâce aux corrélations avec les variables et l'analyse en composantes principales permettent de renommer ces clusters avec les noms respectifs suivant :

Cluster 4 : Le groupe des « Outsiders »

Cluster 5 : Le groupe des « Favoris »

Cluster 6 : Le « Super-Favori »