

# Visualisation de données sur la NBA

Aissam KADRI

Telecom Nancy  
Mars 2024

# 1 Introduction

Ce projet s'inscrit dans le module de Visualisation de Données, enseigné à Telecom Nancy. Le but est de produire une visualisation pertinente d'une base de données, permettant de répondre à un questionnaire personnel.

Dans notre cas, nous travaillons sur les statistiques des joueurs de la NBA sur les années 2021-22. Le choix de cette base de données vient d'un intérêt personnel de certains membres du groupe pour ce sport.

Cette base est composée de deux tables, l'une pour la saison régulière, et l'autre sur les playoffs. En NBA, il existe en effet deux compétitions. Une première phase consiste à déterminer quelles sont les meilleures équipes, en éliminant les moins bonnes (phase régulière). La seconde consiste à faire s'affronter les meilleures équipes, pour déterminer laquelle est la meilleure (phase de playoffs). En conséquence, la table de phase régulière est bien plus consistante que celle des playoffs. Cependant, celle des playoffs contient des informations plus représentatives des performances attendues pendant un match.

Cette base contient de nombreuses données sur les performances des joueurs. Par exemple le taux de tirs à 2 ou 3 points réussis, le nombre de passes décisives, l'âge, le poste des joueurs ... etc. Tout cela permet de réaliser des statistiques sur les performances moyennes des joueurs.

Possédant ces informations, on peut se poser plusieurs questions. Les premières questions qui semblent les plus évidentes consistent à se demander quels joueurs ont été les meilleurs pendant la saison.

Ensuite, on peut se questionner sur les profils types des joueurs. En effet, il existe au basket 5 postes (Pivot, Meneur, Arrière, Ailier, et Ailier Fort), correspondant chacun à une fonction pendant un match. Ainsi, un ailier aura un comportement plus défensif qu'un meneur ou un pivot. On peut ainsi se demander quels sont les profils types de chaque poste.

Une fois les profils type des postes établis, on peut se demander quel poste correspond le mieux à chaque joueur. On peut ainsi déterminer si un joueur est au poste qui lui correspond le plus, ou si celui-ci serait plus adapté à un autre poste.

Enfin, on peut se demander si certains facteurs influent sur les performances des joueurs. Les premiers critères qui viennent en tête sont l'âge et le temps de jeu. On peut se figurer qu'un joueur fatigue lorsqu'il joue depuis suffisamment de temps. Cependant, il existe peut-être d'autres critères pouvant influencer les performances.

La visualisation sur ces données peut être intéressante, notamment pour les coachs des joueurs. En effet, cela permet d'extraire des informations sur ce que doivent être les objectifs des joueurs en fonction de leur style de jeu, et sur les facteurs pouvant altérer leurs performances.

## 2 Prétraitements

Nous cherchons à inférer des informations sur les caractéristiques des joueurs en fonction de leurs statistiques. Pour une meilleure représentativité des résultats, nous avons choisis de travailler majoritairement sur les playoffs, ceux-ci regroupant les meilleurs joueurs. Nous pensons ainsi que ceux-ci représentent mieux les comportements types des joueurs lors d'un match. Nous avons également choisi de ne pas prendre en compte certaines valeurs manquantes dans les données. Les joueurs pour lesquels manquaient certaines données ont ainsi été occultés des visualisations.

Nous avons également isolés certaines caractéristiques intéressantes en fonction de ce que nous voulions tester. En caractéristiques offensives, nous avons par exemple pris en compte le taux de fautes par match, le taux de tirs à 2 ou 3 points réussis ... etc. Pour les statistiques défensives, le taux de passes décisives, le nombre de contres par match, le nombre de rebonds défensifs par matches ... etc.

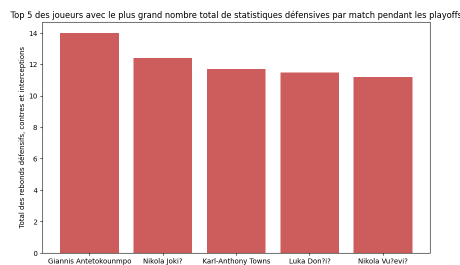
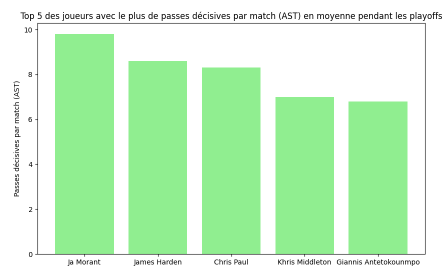
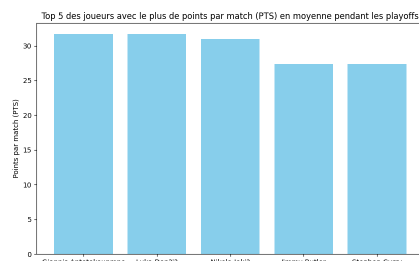
### 3 Quels sont les meilleurs joueurs de la compétition ?

Il existe plusieurs façons d'aborder cette question. En effet, les critères envisagés sont différents, en fonction que l'on value les caractéristiques défensives ou offensives d'un joueur. Nous avons donc décider de produire plusieurs diagramme en fonction des critères que nous jugeons intéressants.

Nous avons envisagé trois métriques pour répondre à cette question :

1. Une basée sur les statistiques offensives, et notamment sur le nombre de points marqués.
2. Une basée sur les statistiques défensives, en combinant le nombre de contres, de rebonds défensifs, et d'interceptions.
3. Une basée sur les statistiques d'équipes, et notamment sur le nombre de passes décisives par match.

Nous avons choisis pour chaque métrique de ne garder que les 5 meilleurs joueurs, afin que les diagrammes restent lisibles. Pour montrer les performances des meilleurs joueurs, on choisi de représenter ces diagrammes par des histogrammes.



### 4 Quels est le profil type d'un joueur sur un poste ?

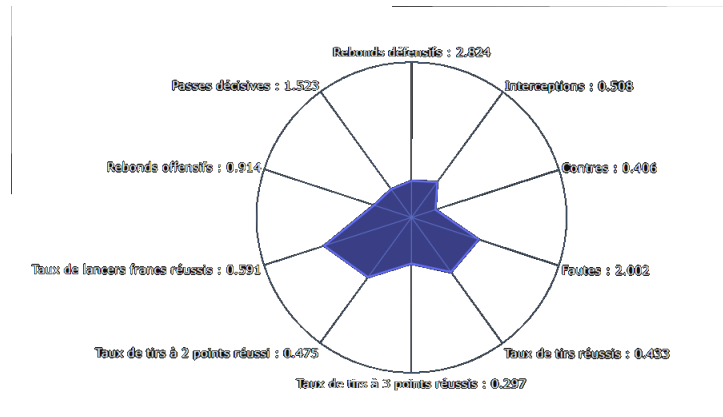
Pour répondre à cette question, plusieurs traitements supplémentaires ont été nécessaires. En effet, il a d'abord fallu séparer les données en fonction des postes de chaque joueur. On obtient ainsi 5 sous-bases de données, chacune correspondant à un poste. Ensuite, il a fallu choisir les caractéristiques que nous jugeons intéressantes pour évaluer

les comportements des joueurs. Nous avons ainsi extrait une dizaine de caractéristiques, permettant d'évaluer les performances offensives et défensives.

Il a ensuite fallu normaliser les valeurs. En effet, les différents attribus ne sont pas représentés de la même façon (pourcentages, mesures, mesures moyennes ... etc.). Cette diversité posait ainsi un problème de représentation sur les graphiques, celles-ci n'ayant pas le même ordre de grandeur. On passe alors par une fonction de normalisation, permettant de ramener chaque valeur à un indice entre 0 et 1 :  $f(X_i) = \frac{X_i - X_{min}}{X_{max} - X_{min}}$ . En conséquence, les valeurs obtenues ne représentent plus les données elles-mêmes, mais un pourcentage de capacité du joueur sur ces critères.

Enfin, afin de représenter les données de manière concise, et vue la diversité des données, nous avons choisi d'utiliser des diagrammes en radar. Ce type de diagramme est particulièrement adapté pour représenter des données multidimensionnelles, ce qui est notre cas.

Nous avons ainsi obtenu 5 diagrammes radar, chacun correspondant à un poste. Par exemple, pour un ailier fort, on obtient le diagramme suivant :



On peut ainsi déceler certaines caractéristiques importantes pour ce poste. Ici, le taux de lancer francs réussis, le nombre de fautes, le taux de tirs réussis ... etc. On peut tracer ce profil pour chaque poste. Nous avons regroupé tous les profils sur une seule image, c.f. le poster final.

## 5 Les joueurs sont-ils adaptés à leur poste ?

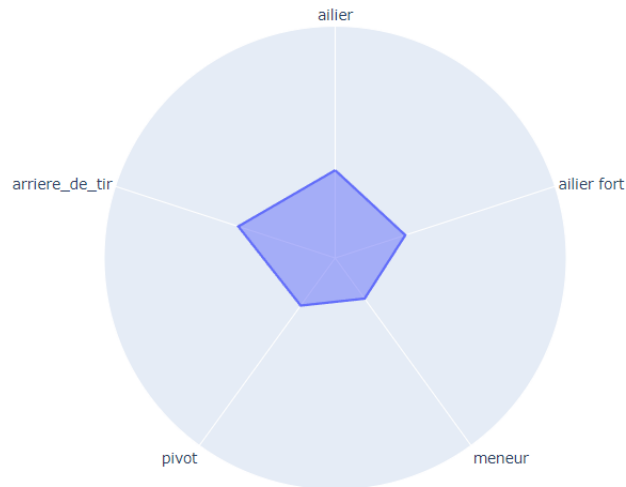
Nous avons maintenant les profils types pour chaque poste. On peut essayer de comparer les caractéristiques des joueurs avec ces profils. Cela permet de déterminer si un joueur est bien adapté au poste qu'il occupe, ou si celui-ci devrait changer de poste.

Pour cela, nous avons créé une fonction permettant d'évaluer la distance entre un joueur et

un poste donné. Pour un poste  $P$ , et un joueur  $X$ , on calcule  $f(X, P) = \sqrt{\sum_{caracteristiques} (X_i - P_i)^2}$

Cette fonction représente la compétence d'un joueur à un poste donné.

On représente le résultat par un diagramme radar, mettant en relation chaque poste avec la distance entre le joueur et le poste. La longueur du radar ne correspond pas cependant à la distance, mais à l'inverse de la distance. En effet, il semble plus parlant de représenter des valeurs plus proches par des aires plus grandes. Par exemple, pour le joueur Steven Adams, on obtient le diagramme suivant :



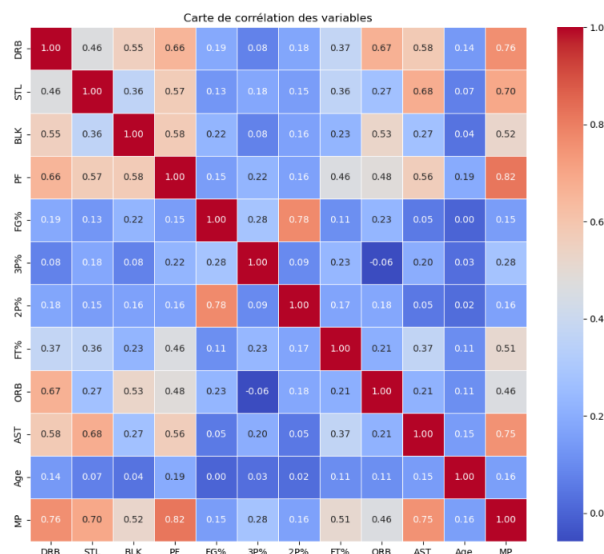
On peut ainsi affirmer que le profil de ce joueur est plus proche d'un arrière ou d'un pivot que d'un ailier. Dans la réalité, ce joueur évoluant au poste d'arrière, on peut affirmer qu'il est dans un poste qui lui correspond bien.

A noter que le diagramme ci-dessus est interactif. Il est possible de choisir le joueur à analyser, et d'obtenir les distances entre son profil et chaque poste.

## 6 Quels facteurs influencent les performances d'un joueur ?

Nous nous sommes enfin demandé quels facteurs pouvaient influencer la performance d'un joueur. Nous pensions que l'âge ou le temps de jeu pourrait faire diminuer les performances d'un joueur. En effet, il semble logique qu'un joueur fatigue après un certain temps de jeu, ou que ces performances soient altérées lorsqu'il vieillit.

Pour vérifier ces affirmations, nous avons mis en corrélation les performances des joueurs en fonction des critères que nous jugeons pertinents. (Ici l'âge et le temps de jeu), en utilisant une carte de chaleur. Nous avons également décidé de corréler les statistiques des joueurs entre elles, pour vérifier la cohérence des valeurs entre elles.



Cette carte de chaleur nous renseigne sur un nombre importants de facteurs. Tout d'abord, on peut remarquer que l'âge du joueur n'est pas si important que nous le pensions initialement. En effet, les coefficients de corrélation ne dépassent pas 0.2, ce qui n'est pas suffisant pour établir une corrélation claire entre l'âge du joueur et ces performances. Cependant, le temps de jeu semble être un critère de corrélation plus intéressant. En effet, on remarque que les corrélations entre le temps de jeu et le nombre de dribbles ou de fautes sont fortement corrélées. Cela paraît logique, puisqu'un temps de jeu plus important augmente la probabilité de commettre des fautes. Enfin, on peut remarquer que les caractéristiques d'un joueur sont assez peu corrélées entre elles. En effet, le taux de tirs à 3 points et le taux de tirs à 2 points réussis sont peu corrélés par exemple. Cela montre qu'un joueur bon au tir à 3 points ne l'est pas forcément pour les tirs à 2 points.

## 7 Conclusion

Le jeu de données fourni nous a permis d'extraire des caractéristiques important pour analyser les performances de chaque joueurs. Nous avons ainsi pu déterminer quels profils chaque joueur devait atteindre en fonction du poste qu'il occupe. Ces informations permettent également de déterminer si un joueur est adapté à son poste, où s'il devrait en changer.

Ensuite, nous avons déterminer quels facteurs pouvaient influencer la performance d'un joueur. Ces facteurs ne sont pas forcément ceux que nous attendions.

En conclusion, les informations que nous avons pu inférer des données fournies permettent de dresser un portrait de la façon dont les joueurs doivent se comporter durant un match. Ces informations peuvent s'avérer très importantes par exemple pour un coach, afin d'adapter les placements des joueurs dans son équipes, ou de développer les compétences de ses joueurs.