



@wesportfr.com

Rapport du projet de développement et déploiement d'une application web R-shiny

Prédiction du salaire des joueurs de NBA en accord
avec leurs performances et statistiques sportives

Louis BERTHIER – Adrien RUGGIERO

TABLE DES MATIERES

I. Définition du sujet	3
1. Origines du projet	3
2. Problématique et intérêts du projet	3
II. La provenance des données	4
III. Notre application web R-shiny	4
1. Onglet n°1 : Comparaison entre les joueurs (cf. Annexe 1)	4
2. Onglet n°2 : Comparaison entre les équipes (cf. Annexe 2)	5
3. Onglet n°3 : Description des données (cf. Annexe 3)	5
4. Onglet n°4 : Présentation des développeurs (cf. Annexe 4)	5
IV. Préconisations	6
V. Annexes	8
1. Annexe n°1	8
2. Annexe n°2	8
3. Annexe n°3	9
4. Annexe n°4	9

I. Définition du sujet

1. Origines du projet

Le basketball, inventé par James Naismith en 1891 devient rapidement de plus en plus populaire en Amérique du Nord. C'est en 1949 que naît la National Basketball Association (NBA), qui aujourd'hui constitue **l'une des quatre ligues majeures** du sport américain. Nous avons tous les 2 un réel intérêt pour le sport et plus particulièrement pour la **NBA**. C'est pourquoi nous avons décidé **d'étudier les salaires des joueurs en fonction de leurs performances sportives**.

En voyant les salaires astronomiques de nombreux joueurs, nous nous étions donc demandés pour le projet de collecte et de stockage des données : **dans quelles mesures les performances sportives des joueurs affectent-elles le salaire de ces derniers ?**

2. Problématique et intérêts du projet

Malheureusement, à la suite de la crise sanitaire, les saisons ont été **perturbées**, on a donc dû se ramener à une **année dite « classique »**. C'est pourquoi nous étudierons **la saison régulière de l'année 2018-2019**.

Les données relatives à la NBA sont **libres d'accès** et retrouvables sur de nombreux sites internet. L'objectif à travers ce projet est de pouvoir **prédire le salaire d'un joueur de NBA en fonction de ses performances** au cours de la saison. Ainsi, un responsable marketing au sein d'une équipe peut essayer de **maximiser les gains et la rentabilité des échanges** de joueurs au cours de chaque nouvelle saison.

En parallèle, cela permettrait par exemple d'adapter les salaires des nouveaux joueurs qui font une première saison fulgurante, que l'on qualifie de « **Rookies** » et dont la valeur « marchande » peut **exploser** au cours de la saison. En parallèle, cela permet de vérifier la valeur des anciens joueurs mondialement connus, que l'on qualifie de « **Stars** » et dont les performances peuvent **chuter** d'une saison à l'autre.

II. La provenance des données

Les données sont issues du projet de « *Collecte et stockage des données* ». Voici un aperçu de notre jeu de données. Plus précisément, il s'agit d'un **fichier csv**.

PLAYER_NAME	PLAYER_ID	NICKNAME	TEAM_ID	TEAM_ABBREVIATION	AGE	GP	W	L	W_PCT	MIN	FGM	FGA	FG_PCT
Aaron Gordon	203932	Aaron	1610612753	ORL	23.0	78	40	38	0.513	33.8	6.0	13.4	0.449
Aaron Holiday	1628988	Aaron	1610612754	IND	22.0	50	31	19	0.62	12.9	2.1	5.2	0.401
Abdel Nader	1627846	Abdel	1610612760	OKC	25.0	61	38	23	0.623	11.4	1.5	3.5	0.423
Al Horford	201143	Al	1610612738	BOS	33.0	68	41	27	0.603	29.0	5.7	10.6	0.535
Al-Farouq Aminu	202329	Al-Farouq	1610612757	POR	28.0	81	52	29	0.642	28.3	3.2	7.3	0.433
Alan Williams	1626210	Alan	1610612751	BKN	26.0	5	1	4	0.2	5.1	1.6	2.6	0.615
Alec Burks	202692	Alec	1610612758	SAC	27.0	64	19	45	0.297	21.5	3.0	7.4	0.405
Alex Abrines	203518	Alex	1610612760	OKC	25.0	31	21	10	0.677	19.0	1.8	5.1	0.357
Alex Caruso	1627936	Alex	1610612747	LAL	25.0	25	8	17	0.32	21.2	3.1	6.9	0.445
Alex Len	203458	Alex	1610612737	ATL	26.0	77	28	49	0.364	20.1	4.2	8.4	0.494
Alex Poythress	1627816	Alex	1610612737	ATL	25.0	21	7	14	0.333	14.5	1.9	3.9	0.494
Alfonzo McKinnie	1628035	Alfonzo	1610612744	GSW	26.0	72	53	19	0.736	13.9	1.9	3.8	0.487
Alize Johnson	1628993	Alize	1610612754	IND	23.0	14	9	5	0.643	4.6	0.3	1.1	0.25
Allen Crabbe	203459	Allen	1610612751	BKN	27.0	43	19	24	0.442	26.4	3.2	8.7	0.367
Allonzo Trier	1629019	Allonzo	1610612752	NYK	23.0	64	13	51	0.203	22.8	3.6	8.1	0.448
Amile Jefferson	1628518	Amile	1610612753	ORL	26.0	12	9	3	0.75	5.7	0.8	1.3	0.625
Amir Johnson	101161	Amir	1610612755	PHI	32.0	51	32	19	0.627	10.4	1.5	3.1	0.503
Andre Drummond	203083	Andre	1610612765	DET	25.0	79	40	39	0.506	33.5	7.1	13.3	0.533
Andre Iguodala	2738	Andre	1610612744	GSW	35.0	68	47	21	0.691	23.2	2.2	4.4	0.5
Andre Ingram	201281	Andre	1610612747	LAL	33.0	4	1	3	0.25	3.7	0.0	1.5	0.0
Andrew Bogut	101106	Andrew	1610612744	GSW	34.0	11	8	3	0.727	12.2	1.6	3.3	0.5
Andrew Harrison	1626150	Andrew	1610612740	NOP	24.0	17	6	11	0.353	10.9	0.9	3.1	0.302
Andrew Wiggins	203952	Andrew	1610612750	MIN	24.0	73	31	42	0.425	34.8	6.8	16.6	0.412
Anfernee Simons	1629014	Anfernee	1610612757	POR	20.0	20	17	3	0.85	7.0	1.4	3.2	0.444
Angel Delgado	1629116	Angel	1610612746	LAC	24.0	2	0	2	0.0	7.4	0.5	2.5	0.2
Ante Zizic	1627790	Ante	1610612739	CLE	22.0	59	13	46	0.22	18.3	3.1	5.6	0.553
Anthony Davis	203076	Anthony	1610612740	NOP	26.0	56	25	31	0.446	33.0	9.5	18.3	0.517
Anthony Tolliver	201229	Anthony	1610612750	MIN	34.0	65	27	38	0.415	16.6	1.5	4.0	0.382
Antonio Blakeney	1628469	Antonio	1610612741	CHI	22.0	57	14	43	0.246	14.5	2.9	6.9	0.419

Bien sûr il ne s'agit que d'un extrait, en réalité il y a **beaucoup plus d'observations et de caractéristiques**.

III. Notre application web R-shiny

Notre application a pour but d'afficher et d'étudier **les statistiques descriptives et sportives rattachées aux joueurs et aux équipes**. Cela permettra tout d'abord à l'utilisateur d'avoir une idée des capacités des entités concernées avec en plus la possibilité de **les comparer** pour se faire une idée du niveau relatif entre les joueurs et du prestige d'une équipe. A la suite, une présentation du **dashboard** que nous avons implémenté.

1. Onglet n°1 : Comparaison entre les joueurs (cf. [Annexe 1](#))

L'idée derrière cette première page est de pouvoir **confronter deux joueurs** d'un point de vue statistique. Bien sûr, il est possible de ne sélectionner **qu'un seul joueur** pour étudier uniquement ses performances. On y retrouve plusieurs graphiques :

- Un diagramme de Kiviat pour **les performances** au cours d'un match établies **avec des nombres entiers ou décimaux**
- Un second diagramme de Kiviat pour **les performances établies avec des pourcentages**. En effet, il y avait un **problème de mise à l'échelle** avec les données du graphique précédent, nous avons donc préféré les distinguer pour **faciliter la lecture et donc l'analyse**

- Un histogramme qui **affiche le salaire** associé à chacun des joueurs en million de dollars (\$)
- Des boîtes à moustaches pour **positionner les joueurs vis-à-vis de la globalité de notre jeu de données**. Ce diagramme nous fournit des **statistiques générales** et expose donc un **point de vue global** avec l'ensemble des joueurs et non un point de vue local avec seulement un ou deux joueurs

2. Onglet n°2 : Comparaison entre les équipes (cf. [Annexe 2](#))

Maintenant que nous nous sommes intéressés aux joueurs à travers la première page, nous avons souhaité nous intéresser à des statistiques **en confrontant les équipes de NBA**. Pour cela nous prenons **les statistiques des joueurs rattachés à une même équipe** et on fait **la moyenne sur chacune des caractéristiques** pour avoir une même base de comparaison.

Notre graphique est un diagramme à bulles qui va **positionner les équipes** dans un espace **selon deux caractéristiques**. En parallèle des deux caractéristiques, on donne la possibilité à l'utilisateur de **filtrer les équipes** en imposant **un nombre de victoires moyen ainsi que l'âge moyen au sein des équipes**. L'ensemble des équipes n'appartenant pas aux conditions de ces filtres ne seront donc **pas considérées pour l'affichage**.

3. Onglet n°3 : Description des données (cf. [Annexe 3](#))

Cette page ne propose pas beaucoup de fonctionnalités. Elle permet simplement à l'utilisateur de **prendre connaissance du langage et du lexique utilisé** dans le monde de la NBA pour **définir et catégoriser les performances sportives** lors des différentes saisons. En réalité, il s'agit d'un simple **glossaire**.

4. Onglet n°4 : Présentation des développeurs (cf. [Annexe 4](#))

Ce dernier onglet représente une **page de contact**. En effet, elle permet de **présenter les deux étudiants** ayant travaillé sur ce projet **avec une photo** de chacun ainsi que **leur adresse mail** respective pour entrer en contact si nécessaire. Également, le **profil LinkedIn** a été mis à disposition à travers **un simple clic**.

On propose aussi **un lien OneDrive** afin de **regrouper l'ensemble des éléments** qui constituent ce projet, à savoir :

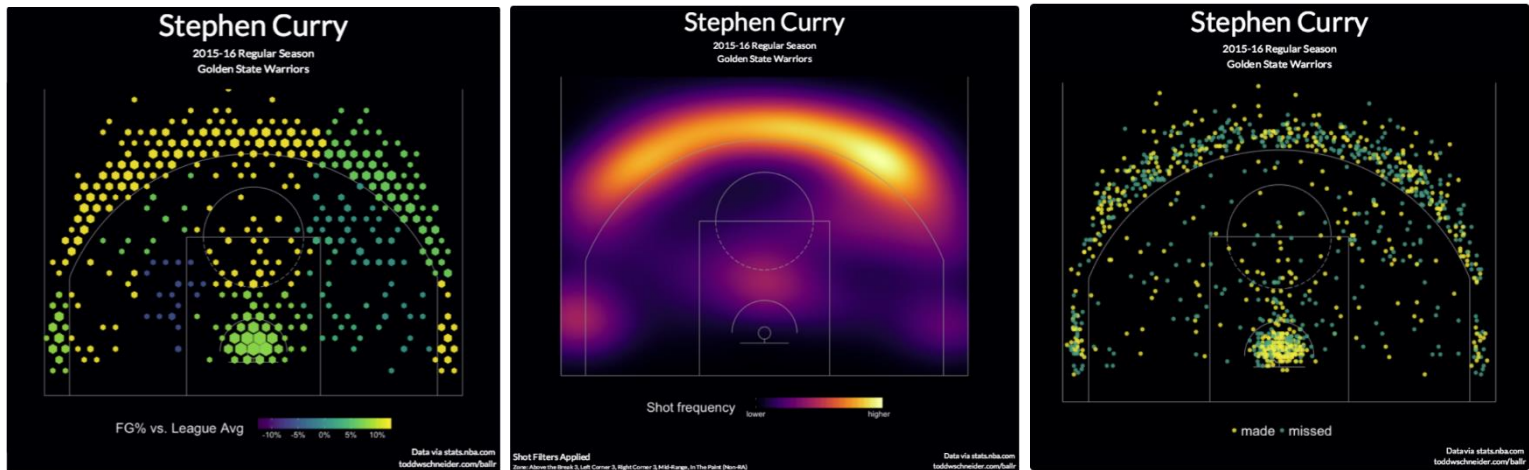
- **Ce rapport** au format pdf
- **Le fichier csv** contenant les données à exploiter
- L'ensemble des **fichiers R** : ui, server et global

IV. Préconisations

Voici quelques pistes afin de **compléter et d'améliorer notre projet** :

- Proposer une partie d'**apprentissage automatique avec apprentissage dit supervisé** : nous avons des **données labélisées** et un **problème de régression** (estimation du salaire). On peut notamment utiliser la **régression linéaire** comme modèle d'apprentissage. Pour cela on aurait pu **restreindre légèrement le nombre de caractéristiques** en se focalisant seulement sur certaines comme le nombre de victoires, le nombre de points, de tirs réussis et tentés, etc...
- **Diversifier l'analyse** et les possibilités en augmentant la taille de notre jeu de données. Pour cela on aurait également pu **recupérer des données de saisons différentes** de celle-ci à savoir 2018-2019. Cela nous aurait notamment permis **d'étudier et de comparer l'évolution des joueurs et des équipes**
- Récupérer un jeu de données contenant **la position des joueurs sur le terrain** lors de certaines interventions (les tirs, les passes, les contres, etc...) pour avoir une **carte visuelle** ainsi qu'une idée de **la répartition et du positionnement des joueurs** lors de ces actions

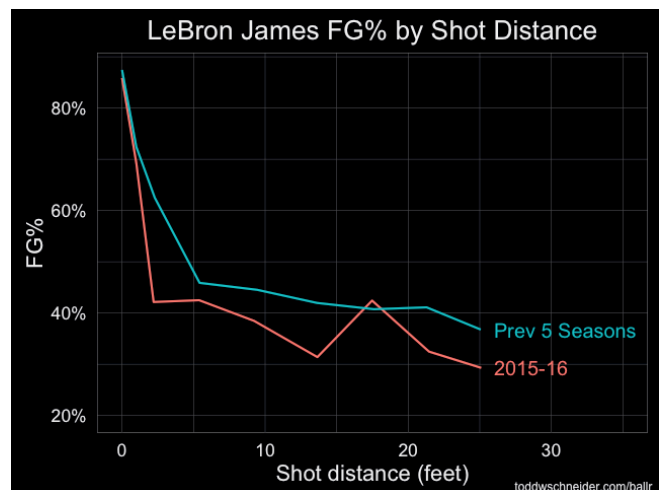
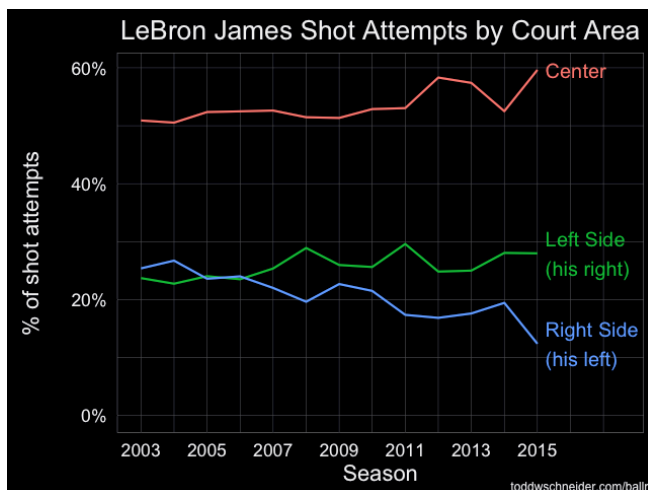
Voici des exemples d'amélioration :



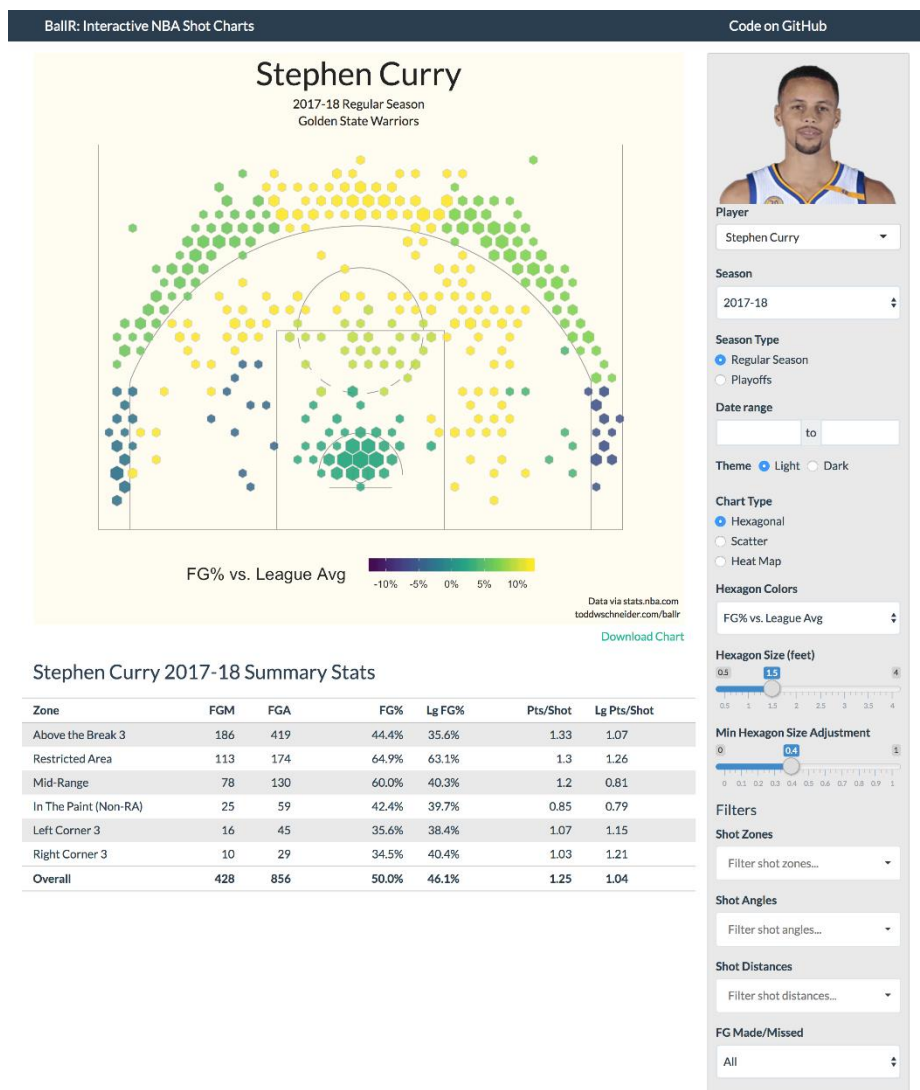
<https://toddwschneider.com/posts/ballr-interactive-nba-shot-charts-with-r-and-shiny/>

- **Améliorer l'interface graphique** en proposant un **visuel plus agréable et plus complet** au niveau des fonctionnalités

Voici quelques illustrations pouvant potentiellement valoriser notre dashboard :



<https://toddwschneider.com/posts/ballr-interactive-nba-shot-charts-with-r-and-shiny/>



<https://github.com/toddwschneider/ballr>

V. Annexes

1. Annexe n°1

BERTHIER - RUGGIERO - NBA Stats Comparaison entre les joueurs Comparaison entre les équipes Description des abréviations Développeurs

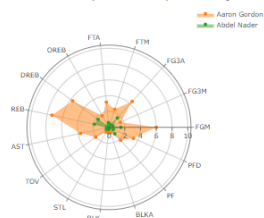
Choix des joueurs

Sélectionner le 1er joueur:
Aaron Gordon - ORL

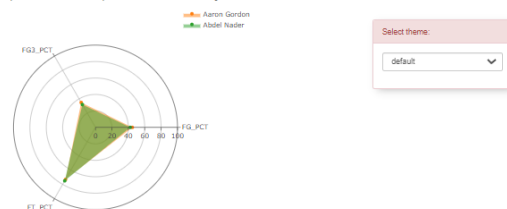
Sélectionner le 2nd joueur:
Abdel Nader - OKC

Filter

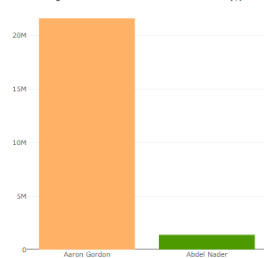
Confrontation des performances sportives des joueurs



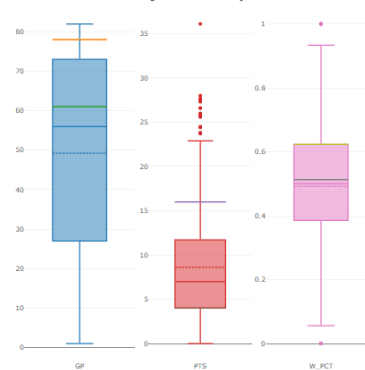
Comparaison en % des performances entre les joueurs



Affichage des salaires en million de dollars (\$)



Evaluation générale des deux joueurs



2. Annexe n°2

BERTHIER - RUGGIERO - NBA Stats Comparaison entre les joueurs Comparaison entre les équipes Description des abréviations Développeurs

Options de filtrage

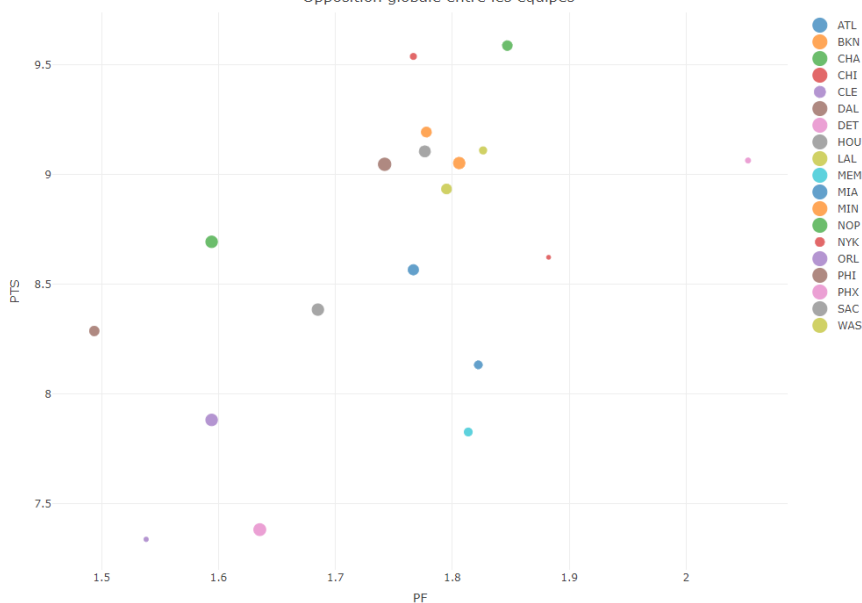
Age moyen au sein de l'équipe :
19 23 35

Nombre moyen de victoires par équipe :
10 30 40

Rechercher

Indication : Chaque bulle a une taille différente qui correspond au nombre moyen de victoires de l'équipe concernée

Opposition globale entre les équipes



3. Annexe n°3

BERTHIER - RUGGIERO - NBA Stats Comparaison entre les joueurs Comparaison entre les équipes Description des abréviations Développeurs

Voici le glossaire à utiliser afin de comprendre les abréviations présentes au sein des différents graphiques :

Précision : Ces données représentent les statistiques en moyenne par match sur l'année

1. Description générale

GP : Nombre de matchs joués | W : Nombre de victoires | L : Nombre de défaites | MIN : Temps joué en minutes

W_PCT : Pourcentage de victoires | FG_PCT : Pourcentage de paniers marqués | FG3_PCT : Pourcentage de paniers marqués à 3 points | FT_PCT : Pourcentage de lancers francs marqués

FGM : Nombre de paniers marqués | FGA : Nombre de tentatives de paniers | FGM3 : Nombre de paniers marqués à 3 points | FGA3 : Nombre de tentatives de paniers à 3 points | FTM : Nombre de lancers francs marqués | FTA : Nombre de tentatives de lancers francs

REB : Nombre total de rebonds | OREB : Nombre de rebonds offensifs | DREB : Nombre de rebonds défensifs

AST : Nombre de passes décisives | TOV : Nombre de pertes de balles | STL : Nombre d'interceptions

BLK : Nombre de contres | BLKA : Nombre de tentatives de contres | PF : Nombre de fautes personnelles | PFD : Nombre de fautes personnelles défensives

PTS : Nombre de points marqués | NBA_FANTASY_PTS : Score fantasiste qui prend en compte des points différents de ceux marqués

Salair : Salaire du joueur en million de dollars (\$)

2. Description par rang

GP_RANK : Rang vis-à-vis du nombre de matchs joués | W_RANK : Rang vis-à-vis du nombre de victoires | L_RANK : Rang vis-à-vis du nombre de défaites | MIN_RANK : Rang vis-à-vis du temps joué

W_PCT_RANK : Rang vis-à-vis du pourcentage de victoires | FG_PCT_RANK : Rang vis-à-vis du pourcentage de paniers marqués | FG3_PCT_RANK : Rang vis-à-vis du pourcentage de paniers marqués à 3 points | FT_PCT_RANK : Rang vis-à-vis du pourcentage de lancers francs marqués

FGM_RANK : Rang vis-à-vis du nombre de paniers marqués | FGA_RANK : Rang vis-à-vis du nombre de tentatives de paniers | FGM3_RANK : Rang vis-à-vis du nombre de paniers marqués à 3 points | FGA3_RANK : Rang vis-à-vis du nombre de tentatives de paniers à 3 points | FTM_RANK : Rang vis-à-vis du nombre de lancers francs marqués | FTA_RANK : Rang vis-à-vis du nombre de tentatives de lancers francs

REB_RANK : Rang vis-à-vis du nombre total de rebonds | OREB_RANK : Rang vis-à-vis du nombre de rebonds offensifs | DREB_RANK : Rang vis-à-vis du nombre de rebonds défensifs

AST_RANK : Rang vis-à-vis du nombre de passes décisives | TOV_RANK : Rang vis-à-vis du nombre de pertes de balles | STL_RANK : Rang vis-à-vis du nombre d'interceptions

BLK_RANK : Rang vis-à-vis du nombre de contres | BLKA_RANK : Rang vis-à-vis du nombre de tentatives de contres | PF_RANK : Rang vis-à-vis du nombre de fautes personnelles | PFD_RANK : Rang vis-à-vis du nombre de fautes personnelles défensives

PTS_RANK : Rang vis-à-vis du nombre de points marqués | NBA_FANTASY_PTS_RANK : Rang vis-à-vis du score fantasiste

4. Annexe n°4

BERTHIER - RUGGIERO - NBA Stats Comparaison entre les joueurs Comparaison entre les équipes Description des abréviations Développeurs

Louis BERTHIER

Vous pouvez me contacter à l'adresse mail suivante : louis.berthier@mines-ales.org

Vous pouvez retrouver mon profil LinkedIn [ici](#)



Adrien RUGGIERO

Vous pouvez me contacter à l'adresse mail suivante : adrien.ruggiero@mines-ales.org

Vous pouvez retrouver mon profil LinkedIn [ici](#)



Notre projet est disponible juste [ici](#). Vous y trouverez notre rapport ainsi que notre code.