

**Projet data mining**

**Application de Clustering NBA**

**Rapport**

**Réalisé par:**

**Yanis Chaibi**  
**Neil Montant**  
**Amina Benzina**

<b>Application de Clustering NBA.....</b>	<b>1</b>
<b>Introduction:.....</b>	<b>3</b>
Objectifs du Projet.....	3
Fonctionnalités de l'Application.....	4
Technologies Utilisées.....	4
<b>Choix des Graphes dans la Partie Statistique Exploratoire.....</b>	<b>5</b>
Onglet "Posts" : Visualisation des Statistiques par Position.....	5
Caractéristiques Principales.....	5
Raisons du Choix.....	5
Onglet "Stats" : Visualisation des Statistiques exploratoire:.....	6
Raisons du Choix.....	6
<b>Choix des Méthodes pour la Partie Modélisation.....</b>	<b>7</b>
Méthodes de Classification.....	7
Conclusion : Choix de la SVM.....	8
<b>Choix des Graphes pour la Visualisation des Performances du Modèle.....</b>	<b>8</b>
Saisie Manuelle pour la Prédiction.....	9
<b>Interpretation des graphiques:.....</b>	<b>9</b>
Analyse Comparative entre les statiques des joueurs centre et des joueurs Shooting Guards.....	9
Analyse des Graphiques des Statistiques par Position.....	11
1. Average Games Played by Position.....	11
2. Average Field Goals per Game by Position.....	11
Interprétation de la Heatmap de Corrélation des Statistiques des Joueurs.....	12
Analyse des Corrélations Clés.....	13
Interprétation de la Matrice de Confusion.....	14
Analyse des Résultats.....	14
Interprétation du Graphique des Prédictions Correctes vs Incorrectes par Position.....	16
Analyse des Résultats.....	16
<b>Conclusion.....</b>	<b>17</b>
Objectifs et Réalisations.....	17
Insights Clés.....	18
Recommandations et Améliorations Futures.....	18

# Introduction

Bienvenue dans notre projet de clustering NBA, une application interactive développée avec Streamlit. Cette application a pour objectif principal d'analyser les performances des joueurs de la NBA afin de déterminer si un joueur est à la bonne position sur le terrain. De plus, elle permet de prédire la position optimale d'un joueur à partir de ses données de performance.

## Objectifs du projet

L'objectif principal de ce projet est de fournir un outil visuel et analytique qui permet aux utilisateurs d'évaluer et de prédire la position optimale des joueurs de basketball sur le terrain. En utilisant des données statistiques et des algorithmes de machine learning, l'application aide à identifier les meilleures positions pour chaque joueur, facilitant ainsi les décisions stratégiques pour les entraîneurs et les analystes.

## Fonctionnalités de l'application

### Chargement et prétraitement des données :

- Importation des données de performance des joueurs depuis diverses sources.
- Nettoyage et normalisation des données pour assurer une analyse cohérente.

### Analyse exploratoire des données:

- Visualisation des statistiques de base des joueurs.
- Graphiques interactifs pour explorer les tendances et les distributions des données.

### Clustering des joueurs :

- Application de différentes méthodes de clustering (K-means, DBSCAN, etc.) pour regrouper les joueurs selon leurs caractéristiques de performance.
- Visualisation des clusters à l'aide de graphiques 2D et 3D.

### Détermination et prédiction de la position :

- Analyse des clusters pour déterminer la position optimale de chaque joueur.
- Utilisation d'algorithmes de classification pour prédire la position d'un joueur en fonction de ses données de performance.

### Interface utilisateur interactive :

- Utilisation de Streamlit pour créer une interface utilisateur intuitive et interactive.
- Widgets pour filtrer et ajuster les paramètres de clustering et de prédiction en temps réel.

## Technologies utilisées

- **Streamlit** : Pour la création de l'interface web interactive.
- **Python** : Pour l'analyse des données et le développement des algorithmes de clustering et de classification.
- **Pandas, NumPy** : Pour la manipulation et l'analyse des données.
- **Scikit-learn** : Pour les algorithmes de machine learning.
- **Matplotlib, Seaborn** : Pour la visualisation des données.

## Choix des graphes dans la partie statistique exploratoire

Dans la partie statistique exploratoire de l'application de clustering NBA, il est crucial de choisir des graphes appropriés pour visualiser les données et extraire des insights significatifs. La partie statistique exploratoire se divise sur deux vue: Postes et stats:

### Onglet "Posts" : Visualisation des statistiques par position

L'onglet "Stats" de notre application de clustering NBA est conçu pour fournir une vue détaillée des performances moyennes des joueurs en fonction de leur position sur le terrain. Cette section permet aux utilisateurs de sélectionner une position spécifique et de visualiser les statistiques correspondantes de manière claire et concise.

## Caractéristiques principales

### Sélection de la position :

- **Fonctionnalité** : Un menu déroulant permet aux utilisateurs de sélectionner la position d'intérêt (par exemple, "Point Guard", "Shooting Guard", etc.).
- **Pourquoi** : Cela permet une exploration ciblée des données, facilitant la comparaison des performances moyennes entre différentes positions.

### Tableau statistique :

- **Fonctionnalité** : Un tableau interactif affiche les statistiques moyennes des joueurs pour la position sélectionnée.
- **Pourquoi** : Le tableau présente des statistiques clés telles que les points marqués par match, les tentatives de tir, les pourcentages de réussite, les

rebonds, les passes décisives, etc., offrant une vue d'ensemble complète des performances des joueurs à une position donnée.

### Raisons du choix

- **Comparaison facile** : Le tableau permet aux utilisateurs de comparer rapidement et facilement les statistiques moyennes des joueurs par position, aidant ainsi à identifier les forces et les faiblesses typiques des joueurs à chaque position.
- **Clarté et Simplicité** : La présentation sous forme de tableau est claire et facile à comprendre, même pour les utilisateurs sans expertise technique ou statistique.
- **Support à la décision** : En fournissant des statistiques détaillées, cet onglet aide les entraîneurs et les analystes à prendre des décisions éclairées concernant la répartition des joueurs sur le terrain et à évaluer si un joueur occupe la position optimale en fonction de ses performances.
- **Exploration interactive** : L'interface interactive permet aux utilisateurs d'explorer les données de manière dynamique, en changeant facilement de position pour voir comment les statistiques varient.

## Onglet "Stats" : Visualisation des statistiques exploratoire:

Dans cette partie de l'application de clustering NBA, l'onglet "Stats" permet une analyse approfondie des statistiques des joueurs en fonction de différentes métriques et positions. Voici une explication détaillée des graphes utilisés et leur pertinence dans l'analyse des données des joueurs de la NBA.

### Graphique en Barres Horizontales :

Un menu déroulant permet aux utilisateurs de sélectionner la statistique d'intérêt (par exemple, "Points per game", "Assists per game", etc.). Cela permet aux utilisateurs de cibler précisément la statistique qu'ils souhaitent analyser, facilitant une exploration plus détaillée des performances des joueurs

- **Utilisation** : Un graphique en barres horizontales affiche les moyennes des statistiques sélectionnées pour chaque position.
- **Pourquoi** : Les barres horizontales sont idéales pour comparer les valeurs moyennes entre différentes catégories (ici, les positions). Elles offrent une visualisation claire et facile à lire des performances des joueurs par position, permettant de voir rapidement quelles positions excèdent ou sont en dessous des autres en termes de performance pour la statistique sélectionnée.

### Carte de Chaleur (Heatmap) des Statistiques de Corrélation :

- **Utilisation** : Une carte de chaleur affiche les corrélations entre différentes statistiques des joueurs.
- **Pourquoi** : La heatmap est un outil puissant pour visualiser les corrélations entre plusieurs variables. Elle aide à identifier les relations linéaires fortes ou faibles entre les différentes statistiques de performance, fournissant un

aperçu des variables qui peuvent être prédictives les unes des autres ou qui varient ensemble.

### Raisons du Choix

- **Comparaison Facile et Rapide** : Le graphique en barres permet une comparaison rapide des performances moyennes des joueurs par position pour la statistique sélectionnée, facilitant ainsi l'identification des positions fortes ou faibles pour cette statistique.
- **Clarté Visuelle** : Les barres horizontales sont intuitives et faciles à lire, même pour les utilisateurs sans formation statistique. Elles offrent une vue claire et ordonnée des données.
- **Insight Profond** : La heatmap des corrélations fournit des insights profonds sur les relations entre les différentes statistiques, ce qui peut être crucial pour les analyses de clustering et de prédiction. Elle aide à comprendre comment les différentes performances des joueurs sont interconnectées.

## Choix des Méthodes pour la Partie Modélisation

Dans la partie modélisation de l'application de clustering NBA, plusieurs méthodes de classification ont été explorées pour prédire la position optimale des joueurs en fonction de leurs performances statistiques. Les méthodes testées incluent K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, et Random Forest. Après une évaluation comparative, nous avons choisi la SVM comme la méthode la plus performante. Voici une analyse détaillée des choix des graphiques et des méthodes utilisées, ainsi que la fonctionnalité de saisie manuelle des prédictions.

### Méthodes de Classification

1. **K-Nearest Neighbors (KNN)** :
  - **Description** : KNN est un algorithme de classification simple et intuitif qui classe un joueur en fonction des positions des K joueurs les plus proches dans l'espace des caractéristiques.
  - **Avantages** : Facile à comprendre et à implémenter. Efficace pour des données avec des structures simples.
  - **Inconvénients** : Peut devenir inefficace avec des jeux de données plus larges ou complexes. Sensible aux valeurs aberrantes et au choix de K.
2. **Support Vector Machine (SVM)** :
  - **Description** : SVM est un algorithme de classification qui trouve l'hyperplan optimal séparant les différentes classes dans un espace de haute dimension.
  - **Avantages** : Utilise des noyaux pour gérer les séparations non linéaires. Maximise les marges entre les classes, améliorant la capacité de

généralisation. Performances supérieures en termes de précision de classification.

- **Inconvénients** : Peut être complexe à interpréter. Le choix du noyau et des paramètres de régularisation est crucial.

### 3. **Decision Tree** :

- **Description** : Un arbre de décision classe les joueurs en divisant récursivement l'espace des caractéristiques en sous-espaces basés sur les valeurs des caractéristiques.
- **Avantages** : Facile à visualiser et à interpréter. Capture les interactions non linéaires entre les caractéristiques.
- **Inconvénients** : Peut sur-apprendre les données d'entraînement, réduisant sa capacité de généralisation.

### 4. **Random Forest** :

- **Description** : Random Forest est un ensemble d'arbres de décision, où chaque arbre est construit à partir d'un sous-échantillon des données.
- **Avantages** : Combine plusieurs arbres de décision pour réduire le risque de sur-apprentissage et améliorer la précision. Fournit des mesures d'importance des variables.
- **Inconvénients** : Plus complexe à interpréter que les arbres de décision individuels. Peut être coûteux en termes de calcul pour les grands ensembles de données.

## Conclusion : Choix de la SVM

Après avoir testé ces différentes méthodes de classification, nous avons choisi la Support Vector Machine (SVM) comme méthode principale pour la modélisation. La SVM s'est avérée la plus efficace pour classer les positions des joueurs en raison de ses performances supérieures, de sa capacité à gérer des données non linéaires et de sa capacité de généralisation.

## Choix des Graphes pour la Visualisation des Performances du Modèle

### Matrice de Confusion :

- **Description** : La matrice de confusion est utilisée pour évaluer les performances du modèle de classification en comparant les prédictions faites par le modèle avec les valeurs réelles. Elle affiche le nombre de prédictions correctes et incorrectes pour chaque classe.
- **Pourquoi** :
  - **Évaluation des Performances** : Permet de visualiser les vraies valeurs positives, les vraies valeurs négatives, les faux positifs et les faux négatifs.

- **Analyse des Erreurs** : Identifie les classes pour lesquelles le modèle se trompe le plus souvent, fournissant des informations sur les améliorations potentielles.
- **Avantages** : Donne une vision claire des performances globales du modèle et des spécificités des erreurs de classification.

#### Prédictions Correctes vs Incorrectes :

- **Description** : Ce graphique en barres compare le nombre de prédictions correctes et incorrectes pour chaque position.
- **Pourquoi** :
  - **Visualisation de la Précision** : Permet de voir rapidement pour quelles positions le modèle fait des prédictions précises et pour quelles positions il se trompe souvent.
  - **Analyse des Performances** : Aide à identifier des tendances ou des anomalies dans les performances du modèle.
- **Avantages** : Fournit une compréhension visuelle et intuitive des points forts et des faiblesses du modèle par position.

#### Saisie Manuelle pour la Prédiction

**Description** : Cette fonctionnalité permet aux utilisateurs d'entrer manuellement les statistiques de performance d'un joueur pour prédire sa position optimale.

##### Pourquoi :

- **Interactivité** : Permet aux utilisateurs d'interagir avec le modèle en entrant des données et en obtenant des prédictions en temps réel.
- **Exploration** : Aide à comprendre comment différentes statistiques influencent la prédiction de la position d'un joueur.
- **Validation** : Les utilisateurs peuvent tester le modèle avec de nouvelles données et vérifier la robustesse des prédictions.

##### Avantages :

- **Personnalisation** : Les utilisateurs peuvent entrer des valeurs spécifiques et obtenir des prédictions adaptées à leurs besoins.
- **Éducation** : Aide les utilisateurs à comprendre l'impact des différentes caractéristiques sur les prédictions de position.

## Interpretation des graphiques:



# Analyse Comparative entre les statiques des joueurs centre et des joueurs Shooting Guards

PostesStats

Select Position

Center

	Average
Field goals per game	4.4199
Field goal attempts per game	7.9156
Field goal percentage	0.5679
3-point field goals per game	0.5062
3-point field goal attempts per game	1.4578
3-point field goal percentage	0.2496
2-point field goals per game	3.9137
2-point field goal attempts per game	6.4555
2-point field goal percentage	0.6107
Effective field goal percentage	0.5980
Free throws per game	1.9536
Free throw attempts per game	2.7081
Free throw percentage	0.7060
Offensive rebounds per game	2.1986
Defensive rebounds per game	5.0152
Total rebounds per game	7.2118
Assists per game	1.9085
Steals per game	0.6464
Blocks per game	0.9910
Turnovers per game	1.3536
Personal fouls per game	2.4114
Points per game	11.2948

PostesStats

Select Position

Shooting Guard

	Average
Field goals per game	4.1802
Field goal attempts per game	9.5006
Field goal percentage	0.4359
3-point field goals per game	1.6665
3-point field goal attempts per game	4.5712
3-point field goal percentage	0.3594
2-point field goals per game	2.5128
2-point field goal attempts per game	4.9339
2-point field goal percentage	0.5070
Effective field goal percentage	0.5267
Free throws per game	1.5735
Free throw attempts per game	1.9406
Free throw percentage	0.7923
Offensive rebounds per game	0.6163
Defensive rebounds per game	2.6150
Total rebounds per game	3.2300
Assists per game	2.6348
Steals per game	0.7863
Blocks per game	0.2974
Turnovers per game	1.3003
Personal fouls per game	1.8026
Points per game	11.5939

- **Centres (Centers)** : Les centres se concentrent principalement sur les tirs à 2 points et les rebonds, avec un haut pourcentage de réussite proche du panier. Ils ont une forte présence défensive, marquant un nombre significatif de rebonds et de blocks.

- **Arrières Shooters (Shooting Guards) :** Les shooting guards sont plus actifs dans les tirs à 3 points, avec une efficacité modérée. Ils participent davantage aux passes décisives et ont un pourcentage de lancers francs plus élevé.

## Analyse des Graphiques des Statistiques par Position



### 1. Average Games Played by Position

Ce graphique montre le nombre moyen de matchs joués par les joueurs à différentes positions.

- **SG-PG (Shooting Guard-Point Guard) :** Les joueurs occupant cette double position jouent en moyenne le plus grand nombre de matchs.
- **C-PF (Center-Power Forward) et C (Center) :** Ces positions sont également parmi les plus fréquentes en termes de matchs joués, suggérant que les joueurs à ces positions ont une grande durabilité ou sont souvent utilisés dans les rotations d'équipe.
- **PG (Point Guard) :** Les meneurs de jeu jouent en moyenne le moins de matchs parmi les positions affichées, ce qui pourrait indiquer une gestion plus stricte du temps de jeu ou des blessures plus fréquentes.

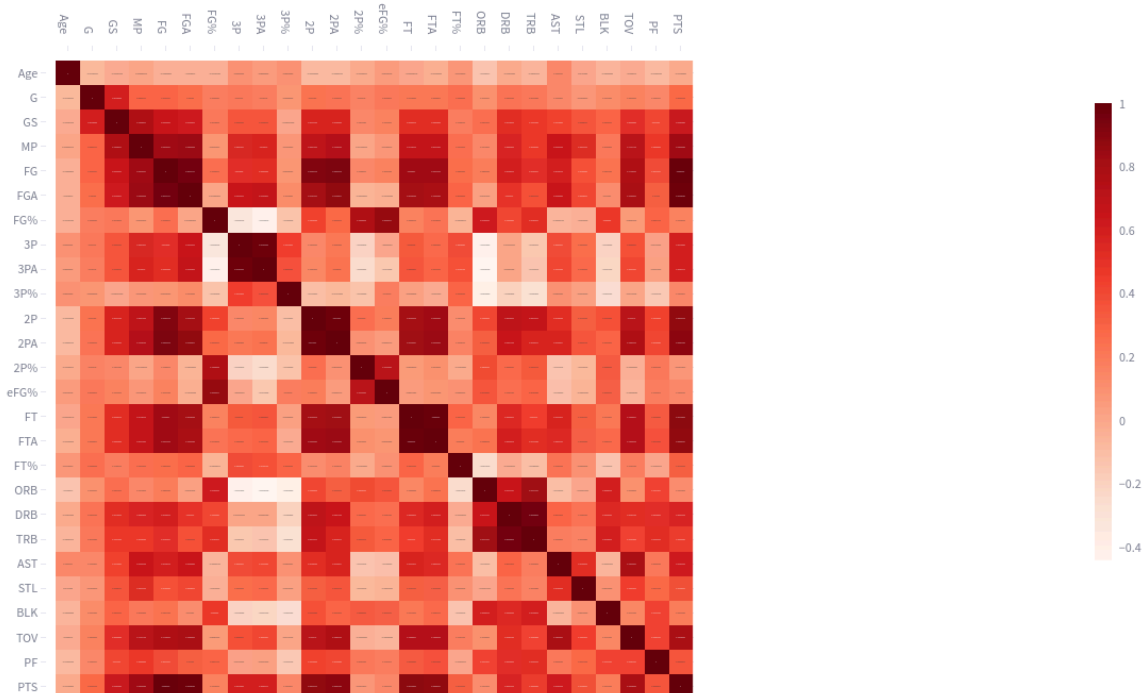
## 2. Average Field Goals per Game by Position

Ce graphique montre le nombre moyen de tirs réussis par match pour les joueurs à différentes positions.

- **SG-PG (Shooting Guard-Point Guard)** : Ces joueurs réussissent le plus de tirs par match en moyenne, ce qui souligne leur rôle crucial dans l'attaque.
- **C (Center)** : Les centres réussissent également un grand nombre de tirs par match, ce qui est typique car ils jouent souvent près du panier.
- **PG (Point Guard)** et **PF (Power Forward)** : Ces positions suivent en termes de tirs réussis par match, indiquant leur importance dans la réalisation de points pour l'équipe.
- **SG-SF (Shooting Guard-Small Forward)** : Cette position a le nombre moyen de tirs réussis par match le plus bas parmi ceux affichés, ce qui pourrait indiquer un rôle plus diversifié ou moins axé sur le scoring direct.

Les graphiques montrent que les joueurs occupant les positions de SG-PG jouent le plus grand nombre de matchs et réussissent le plus de tirs par match, indiquant une importance clé dans l'équipe tant en termes de présence sur le terrain que de contribution offensive. Les centres jouent également un nombre élevé de matchs et réussissent de nombreux tirs, ce qui est conforme à leur rôle traditionnel. Les meneurs de jeu, bien qu'ils jouent moins de matchs en moyenne, restent cruciaux pour les réalisations de points. Ces insights peuvent aider les entraîneurs et les analystes à comprendre les contributions typiques des joueurs à différentes positions et à optimiser les rotations et les stratégies de jeu.

# Interprétation de la Heatmap de Corrélation des Statistiques des Joueurs



Cette heatmap de corrélation représente les coefficients de corrélation entre différentes statistiques de performance des joueurs de la NBA. Les valeurs de corrélation varient de -1 à 1, où -1 indique une corrélation négative parfaite, 0 aucune corrélation, et 1 une corrélation positive parfaite. Les nuances de rouge plus foncées représentent des corrélations plus fortes (positives ou négatives).

## Analyse des Corrélations Clés

### 1. Points par Match (PTS) :

- **Forte Corrélation Positive** : avec les Field Goals par Match (FG), les Tentatives de Field Goal (FGA), les 3-Points par Match (3P), et les Lancers Francs par Match (FT). Cela indique que les joueurs qui marquent plus de points tendent également à prendre plus de tirs et à réussir un plus grand nombre de ces tirs.
- **Corrélation Modérée** : avec les Passes Décisives par Match (AST) et les Rebounds Totaux par Match (TRB), suggérant que les joueurs performants en termes de points peuvent aussi contribuer significativement aux passes décisives et aux rebonds.

### 2. Field Goals par Match (FG) et Tentatives de Field Goal (FGA) :

- **Forte Corrélation Positive** : entre FG et FGA, ainsi qu'entre FGA et Points par Match (PTS), ce qui est attendu car plus un joueur tente de tirs, plus il a de chances de marquer des points.
- **Corrélation avec l'Efficacité (FG%)** : La corrélation entre FG% (pourcentage de réussite des tirs) et FG est modérée, indiquant que les

joueurs qui marquent beaucoup de field goals tendent à être relativement efficaces, mais d'autres facteurs peuvent aussi influencer cette relation.

3. **3-Points par Match (3P) et Tentatives de 3-Points (3PA) :**

- **Forte Corrélation Positive** : entre 3P et 3PA, ainsi qu'avec le pourcentage de réussite des tirs à 3 points (3P%). Les joueurs qui prennent plus de tirs à 3 points tendent à en réussir un plus grand nombre.
- **Corrélation avec les Points (PTS)** : Modérée, indiquant que les tirs à 3 points contribuent de manière significative aux points totaux.

4. **Rebounds Totaux par Match (TRB) :**

- **Forte Corrélation Positive** : avec les Rebounds Offensifs (ORB) et Défensifs (DRB), ce qui montre que les joueurs qui captent beaucoup de rebonds totaux sont actifs des deux côtés du terrain.
- **Corrélation avec les Points (PTS)** : Modérée, ce qui peut suggérer que les joueurs bons aux rebonds ont aussi des opportunités de marquer des points après un rebond offensif.

5. **Passes Décisives par Match (AST) :**

- **Corrélation Positive Modérée** : avec les Points (PTS) et les Field Goals par Match (FG), indiquant que les joueurs qui réalisent plus de passes décisives contribuent également directement ou indirectement aux points de l'équipe.

6. **Lancers Francs par Match (FT) et Tentatives de Lancers Francs (FTA) :**

- **Forte Corrélation Positive** : entre FT et FTA, et avec les Points par Match (PTS), montrant que les joueurs qui obtiennent plus de fautes et tentent plus de lancers francs marquent également plus de points.

Cette heatmap de corrélation offre une vue d'ensemble claire des relations entre les différentes statistiques de performance des joueurs de la NBA. Les corrélations fortes entre les tentatives de tirs et les points marqués, ainsi qu'entre les rebonds offensifs, défensifs et totaux, confirment les rôles attendus des joueurs en fonction de leurs positions. Les informations fournies par cette analyse peuvent être utilisées pour identifier les caractéristiques les plus influentes sur les performances globales des joueurs et optimiser les stratégies de jeu.

## Interprétation de la Matrice de Confusion

Actual	C	93	18	0	3	2
	PF	36	54	2	40	17
	PG	2	3	72	8	48
	SF	7	23	8	45	48
	SG	4	5	23	26	81
		C	PF	PG	SF	SG
		Predicted				

Cette matrice de confusion montre les performances du modèle de classification SVM utilisé pour prédire la position des joueurs de la NBA. Elle compare les positions réelles des joueurs (Actual) avec les positions prédites par le modèle (Predicted).

### Analyse des Résultats

- Center (C) :**
  - **Prédictions Correctes** : 93 (cases sur la diagonale)
  - **Prédictions Incorrectes** : 23 (18 PF, 3 SF, 2 SG)
  - **Interprétation** : Le modèle classe correctement la majorité des centres, mais il y a quelques confusions avec les power forwards (PF).
- Power Forward (PF) :**
  - **Prédictions Correctes** : 54
  - **Prédictions Incorrectes** : 95 (36 C, 40 SF, 17 SG, 2 PG)
  - **Interprétation** : Le modèle a plus de difficulté à classer correctement les power forwards, avec des confusions significatives avec les centres (C) et les small forwards (SF).
- Point Guard (PG) :**
  - **Prédictions Correctes** : 72
  - **Prédictions Incorrectes** : 61 (48 SG, 8 SF, 3 PF, 2 C)
  - **Interprétation** : Les point guards sont souvent confondus avec les shooting guards (SG), ce qui pourrait être dû à des similitudes dans les statistiques de performance de ces deux positions.
- Small Forward (SF) :**

- **Prédictions Correctes** : 45
- **Prédictions Incorrectes** : 86 (48 SG, 23 PF, 8 PG, 7 C)
- **Interprétation** : Les small forwards sont fréquemment confondus avec les shooting guards (SG) et les power forwards (PF), indiquant une difficulté à distinguer ces positions.

#### 5. Shooting Guard (SG) :

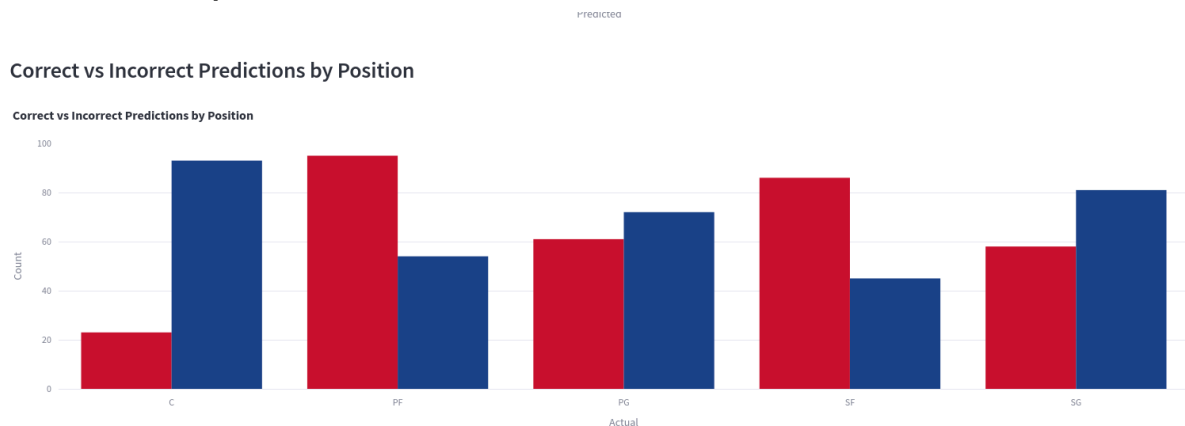
- **Prédictions Correctes** : 81
- **Prédictions Incorrectes** : 58 (26 SF, 23 PG, 5 PF, 4 C)
- **Interprétation** : Bien que le modèle réussisse bien à classer les shooting guards, il existe des confusions notables avec les small forwards (SF) et les point guards (PG).

Cette matrice de confusion révèle plusieurs points clés sur les performances du modèle SVM :

- **Bonne Précision pour les Centers (C) et les Shooting Guards (SG)** : Le modèle parvient bien à classer ces positions, comme en témoigne le nombre élevé de prédictions correctes.
- **Difficultés pour les Power Forwards (PF) et les Small Forwards (SF)** : Il y a une grande quantité de confusions avec d'autres positions, ce qui indique que les caractéristiques statistiques de ces positions peuvent se chevaucher de manière significative.
- **Confusions entre Point Guards (PG) et Shooting Guards (SG)** : Les similitudes dans les tâches et les statistiques de performance entre ces positions expliquent les erreurs de classification.

Ces insights peuvent guider les améliorations futures du modèle, telles que l'ajustement des hyperparamètres, l'intégration de nouvelles caractéristiques, ou l'utilisation de techniques avancées de sélection de caractéristiques pour mieux différencier les positions de jeu.

## Interprétation du Graphique des Prédictions Correctes vs Incorrectes par Position



Ce graphique en barres compare le nombre de prédictions correctes (en bleu) et incorrectes (en rouge) pour chaque position de joueur (Center, Power Forward, Point Guard, Small Forward, Shooting Guard). Il fournit une vue d'ensemble des performances du modèle de classification SVM.

### Analyse des Résultats

1. **Center (C) :**
  - **Prédictions Correctes :** Environ 93
  - **Prédictions Incorrectes :** Environ 23
  - **Interprétation :** Le modèle parvient à classer correctement la majorité des centres, avec un nombre relativement faible d'erreurs. Cela suggère que les caractéristiques des centres sont bien distinctes et bien capturées par le modèle.
2. **Power Forward (PF) :**
  - **Prédictions Correctes :** Environ 54
  - **Prédictions Incorrectes :** Environ 95
  - **Interprétation :** Le modèle éprouve des difficultés à classer correctement les power forwards, avec un nombre élevé de prédictions incorrectes. Cela pourrait indiquer que les statistiques des power forwards se chevauchent avec celles d'autres positions, rendant la classification plus complexe.
3. **Point Guard (PG) :**
  - **Prédictions Correctes :** Environ 72
  - **Prédictions Incorrectes :** Environ 61
  - **Interprétation :** Les point guards sont souvent correctement classifiés, mais il y a encore une proportion significative d'erreurs, principalement en raison de confusions avec les shooting guards (SG).
4. **Small Forward (SF) :**
  - **Prédictions Correctes :** Environ 45
  - **Prédictions Incorrectes :** Environ 86
  - **Interprétation :** Les small forwards sont fréquemment mal classifiés, ce qui pourrait suggérer des similitudes dans les statistiques avec les power forwards (PF) et les shooting guards (SG).
5. **Shooting Guard (SG) :**
  - **Prédictions Correctes :** Environ 81
  - **Prédictions Incorrectes :** Environ 58
  - **Interprétation :** Le modèle parvient à classer correctement un grand nombre de shooting guards, bien que des confusions avec les small forwards (SF) et les point guards (PG) soient présentes.

Ce graphique met en lumière les points forts et les faiblesses du modèle de classification SVM :

- **Bonne Précision pour les Centers et Shooting Guards :** Le modèle réussit bien à classer ces positions, comme le montrent les nombreuses prédictions correctes.
- **Difficultés pour les Power Forwards et Small Forwards :** Ces positions présentent un nombre élevé de prédictions incorrectes, suggérant des défis dans la distinction de leurs caractéristiques statistiques.



- **Confusions entre Point Guards et Shooting Guards :** Les similitudes dans les tâches et les performances statistiques de ces positions expliquent les erreurs de classification.

Ces insights peuvent être utilisés pour améliorer le modèle en explorant des ajustements des hyperparamètres, en intégrant de nouvelles caractéristiques, ou en utilisant des techniques de sélection de caractéristiques pour mieux différencier les positions de jeu.

## Conclusion

Ce projet de clustering NBA, développé avec Streamlit, a permis d'analyser les performances des joueurs et de prédire leur position optimale sur le terrain en utilisant des techniques de machine learning. Voici les points clés et les conclusions tirées de ce projet :

### Objectifs et Réalisations

1. **Analyse Exploratoire des Données :**
  - Nous avons visualisé les statistiques moyennes des joueurs par position, permettant de mieux comprendre les caractéristiques distinctives de chaque rôle sur le terrain.
  - Les graphiques interactifs ont facilité l'exploration des performances des joueurs, en mettant en évidence les tendances et les distributions des données.
2. **Modélisation et Prédiction :**
  - Plusieurs méthodes de classification ont été testées, y compris KNN, SVM, Decision Tree, et Random Forest. Après une évaluation comparative, la SVM a été choisie pour sa précision supérieure dans la prédiction des positions des joueurs.
  - La matrice de confusion et les graphiques des prédictions correctes vs incorrectes ont fourni des insights détaillés sur les performances du modèle, révélant ses points forts et ses faiblesses.
3. **Saisie Manuelle des Prédictions :**
  - La fonctionnalité de saisie manuelle a permis aux utilisateurs d'interagir avec le modèle en entrant des données de performance spécifiques et en obtenant des prédictions en temps réel.
  - Cela a aidé à comprendre l'impact des différentes caractéristiques sur les prédictions de position, offrant une exploration interactive et éducative.

### Insights Clés

- **Centres et Shooting Guards :** Ces positions ont été bien classifiées par le modèle, avec un nombre élevé de prédictions correctes. Les caractéristiques statistiques distinctes de ces rôles ont facilité leur identification.

- **Power Forwards et Small Forwards** : Ces positions ont présenté des défis pour le modèle, avec un nombre élevé de confusions. Les similitudes dans les statistiques de performance avec d'autres positions ont rendu leur classification plus complexe.
- **Point Guards et Shooting Guards** : Les similitudes dans les tâches et les statistiques de performance de ces positions ont conduit à des confusions fréquentes, soulignant la nécessité d'une analyse plus approfondie ou de caractéristiques supplémentaires pour améliorer la différenciation.

## Recommandations et Améliorations Futures

- **Amélioration du Modèle** : Pour réduire les erreurs de classification, il serait bénéfique d'explorer des ajustements des hyperparamètres, d'intégrer de nouvelles caractéristiques pertinentes et d'utiliser des techniques avancées de sélection de caractéristiques.
- **Analyse Complémentaire** : Effectuer une analyse plus approfondie des positions présentant des chevauchements statistiques, comme les power forwards et small forwards, pour identifier des caractéristiques distinctives supplémentaires.
- **Extension des Données** : Enrichir le jeu de données avec des saisons supplémentaires ou des données de ligues différentes pourrait améliorer la robustesse et la généralisation du modèle.

Ce projet a démontré l'efficacité de l'utilisation des techniques de machine learning pour analyser et prédire les positions des joueurs de la NBA. Grâce à des visualisations interactives et à une modélisation rigoureuse, nous avons pu obtenir des insights précieux sur les performances des joueurs et leurs rôles optimaux sur le terrain. Les résultats obtenus offrent une base solide pour des analyses futures et des améliorations continues, contribuant ainsi à une meilleure compréhension et gestion stratégique des talents dans le basketball professionnel.