

PROJET SAS M1 MAS ANALYSE DE DONNÉES DE LA NHL

**PHAM THI MINH NGOC
NGUYEN MANH HUNG
DHENNIN NOLWENN**



PRÉSENTATION DU SUJET:

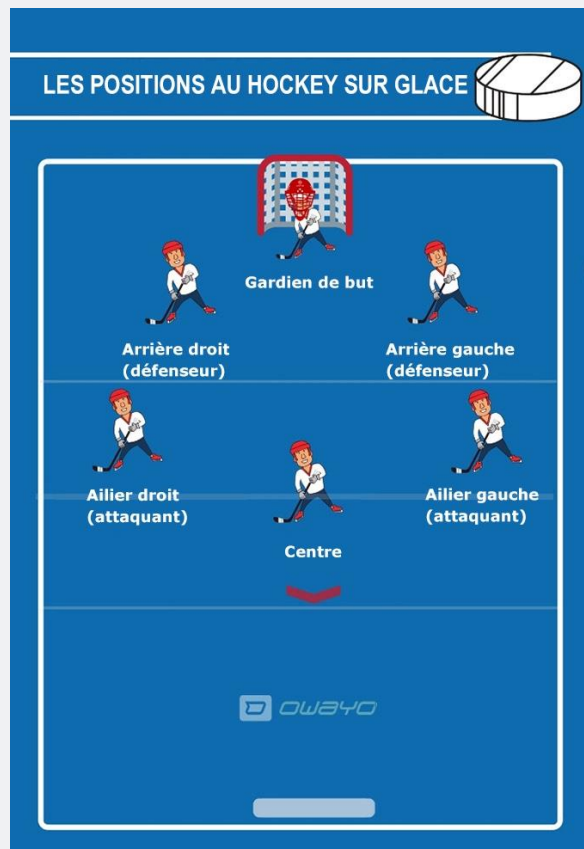
- Nous disposons de bases de données représentant toutes les mesures officielles mesurées pour chaque match dans la National Hockey League au cours des 6 dernières années, les résultats et les statistiques des joueurs des matchs de la NHL, il comprend des détails sur les jeux individuels tels que les tirs, les buts et les arrêts, etc.
- Ces données sont extrêmement réalistes et intéressantes, pour explorer et évaluer les résultats des joueurs nous allons les exploiter dans l'aspect suivant.



Problématique

Quelles sont les différentes positions au hockey sur glace ?

Et surtout, quels sont leurs rôles respectifs ?



- Chaque équipe de hockey sur glace est composée de six joueurs actifs simultanément sur le terrain.
- Les cinq joueurs de champ occupent les positions suivantes : un centre, deux ailiers (droit et gauche), deux défenseurs
- Dans la zone de but se tient le sixième joueur : le gardien de but.

Par exemple: un joueur très offensif peut récolter beaucoup de points, mais si sa capacité à défendre est nulle, son équipe risque de prendre plus de buts qu'elle n'en provoque au final.

Nous allons examiner précisément les caractéristiques attribuées au gardiens de but, aux ailiers, au centre et à la défense des joueurs de la NHL.



Importation des données

Pour répondre à notre problématique, nous avons eu besoin d'importer 5 tables :

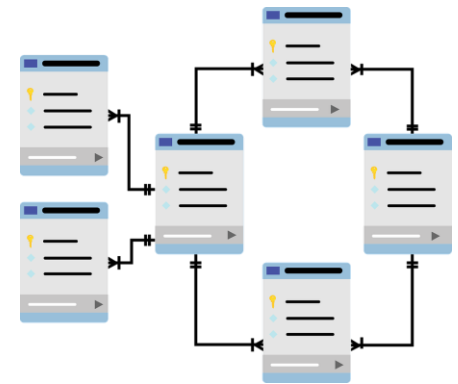
player_info

game_goalie_stats

game_skater_stats

game

game_team_stats



Après avoir nettoyé les données (suppressions de doublons, modification de certains types de variables qui étaient inappropriés, gérer les données manquantes, etc), nous avons pu commencer à les étudier.

Création de tables

1. Indicateurs moyens par joueur des gardiens et hockeyeurs pour les saisons entre 2005 et 2020

- Tout d'abord, nous avons commencé par créer une table permettant d'observer les indicateurs moyens de chaque 'skater', c'est-à-dire de chaque hockeyeur, par saison nommée **ind_moy_skater**. Les joueurs de champ, eux, ont plusieurs tâches. En effet, tout changement dans la situation de jeu requiert de la part des joueurs une certaine polyvalence et un soutien constant à leurs coéquipiers. Cela signifie qu'un défenseur peut très bien attaquer le but adverse en situation offensive et qu'à l'inverse, un attaquant peut prêter main-forte à la défense. Cependant, chaque position se voit attribuée des tâches bien spécifiques.
- Ensuite, nous avons fait la même chose mais pour les gardiens, dans une table nommée **ind_moy_goalie**. Le gardien de but de hockey sur glace défend contre les tentatives de buts de l'adversaire. La tâche principale du gardien de but est de défendre son but en bloquant les palets adverses tout en défendant avec succès ses filets.

```
create table Projet.ind_moy_skater as
select Distinct A.season,
               B.player_id,
               C.primaryPosition,
               avg(timeOnIce) as avg_timeOnIce,
               avg(B.assists) as avg_assists,
               avg(B.goals) as avg_goals,
               avg(B.shots) as avg_shots,
               avg (B.powerPlayAssists) as avg_powerPlayAssists,
               avg (B.faceOffWins) as avg_faceOffWins,
               avg (B.faceoffTaken) as avg_faceoffTaken,
               avg (B.shortHandedGoals) as avg_shortHandedGoals,
               avg (B.shortHandedAssists) as avg_shortHandedAssists,
               avg (B.plusMinus) as avg_plusMinus,
               avg (B.evenTimeOnIce) as avg_evenTimeOnIce,
               avg (B.shortHandedTimeOnIce) as avg_shortHandedTimeOnIce,
               avg (B.powerPlayTimeOnIce) as avg_powerPlayTimeOnIce,
               avg (B.new_blocked) as avg_new_blocked,
               avg (B.new_hits) as avg_new_hits,
               avg (B.new_takeaways) as avg_new_takeaways,
               avg (B.new_giveaways) as avg_new_giveaways
from Projet.game A inner join Projet.GAME_SKATER_STATS B on A.game_id=B.game_id
                  inner join PROJET.PLAYER_INFO C on B.player_id=C.player_id
Where A.season NOT IN ('20002001', '20012002', '20022003', '20032004', '20042005')
Group by A.season, B.player_id
;
```

Création de tables

- À partir des deux tableaux créés ci-dessus, on peut voir qu'il n'y a que 4 variables quantitatives en commun entre les deux tableaux. Donc, nous avons récupéré les indicateurs : *season*, *primaryPosition*, *player_id*, *avg_timeOnIce*, *avg_assists*, *avg_goals*, *avg_shots* pour les saisons entre 2015 et 2020, en créant deux tables : **ind_moy_skater_20152020** (sur les hockeyeurs) et **ind_moy_goalie_20152020** (sur les gardiens).

```
proc sql;
create table Projet.ind_moy_skater_20152020 as
select  season,primaryPosition,player_id,
        avg_timeOnIce, avg_assists,avg_goals,avg_shots
from Projet.ind_moy_skater as A
Where A.season IN ('20152016','20162017',
                  '20172018','20182019','20192020')
;
create table Projet.ind_moy_goalie_20152020 as
select  season,primaryPosition, player_id,
        avg_timeOnIce,avg_assists,avg_goals,avg_shots
from Projet.ind_moy_goalie as A
Where A.season IN ('20152016','20162017',
                  '20172018','20182019','20192020')
;
quit;
data Projet.ind_moy_20152020;
set Projet.ind_moy_skater_20152020 Projet.ind_moy_goalie_20152020;
run
;
```

Création de tables

2. Caractéristiques moyennes pour chaque position des joueurs (centre, ailier droit, ailier gauche, défenseur) sauf le gardien

- Afin de récupérer les caractéristiques moyennes de chaque position associée à chacun des joueurs, nous avons créé une table se nommant **pos_moy**. Dans le jeu, les trois joueurs en position d'attaquant travaillent en étroite collaboration : les deux ailiers (gauche et droit) font passer le palet latéralement au centre, de sorte à ce que dernier puisse tenter un but. Les deux défenseurs de hockey sur glace jouent, tout comme les ailiers, soit sur le côté gauche, soit sur le côté droit du terrain. Leur tâche principale est de protéger leur propre zone de but et de récupérer le palet pour relancer l'attaque sur l'équipe adverse. C'est pourquoi les indicateurs: *avg_goals*, *avg_faceOffWin*, *avg_faceOffTaken*, *avg_shortHandedGoals* des deux défenseurs sont les plus bas. En revanche, les indicateurs: *avg_evenTimeOnIce*, *avg_shortHandedTimeOnIce*, *avg_new_giveaways* sont les plus hauts.

```
create table Projet.pos_moy as
select C.primaryPosition,
       avg(B.assists) as avg_assists,
       avg(B.goals) as avg_goals,
       avg(B.shots) as avg_shots,
       avg (B.powerPlayAssists) as avg_powerPlayAssists,
       avg (B.faceOffWins) as avg_faceOffWins,
       avg (B.faceoffTaken) as avg_faceoffTaken,
       avg (B.shortHandedGoals) as avg_shortHandedGoals,
       avg (B.shortHandedAssists) as avg_shortHandedAssists,
       avg (B.plusMinus) as avg_plusMinus,
       avg (B.evenTimeOnIce) as avg_evenTimeOnIce,
       avg (B.shortHandedTimeOnIce) as avg_shortHandedTimeOnIce,
       avg (B.powerPlayTimeOnIce) as avg_powerPlayTimeOnIce,
       avg (B.new_takeaways) as avg_new_takeaways,
       avg (B.new_giveaways) as avg_new_giveaways
from Projet.game A inner join Projet.GAME_SKATER_STATS B on A.game_id=B.game_id
                  inner join PROJET.PLAYER_INFO C on B.player_id=C.player_id
Group by C.primaryPosition
;
```


Création de tables

3. Nombre de chaque poste par saison à partir de 2015 et caractéristiques moyennes pour chaque position des joueurs

- Pour comparer les différentes saisons, la table **nbe_postes** permettant de récupérer le nombre de joueurs par poste par saison a été créée.

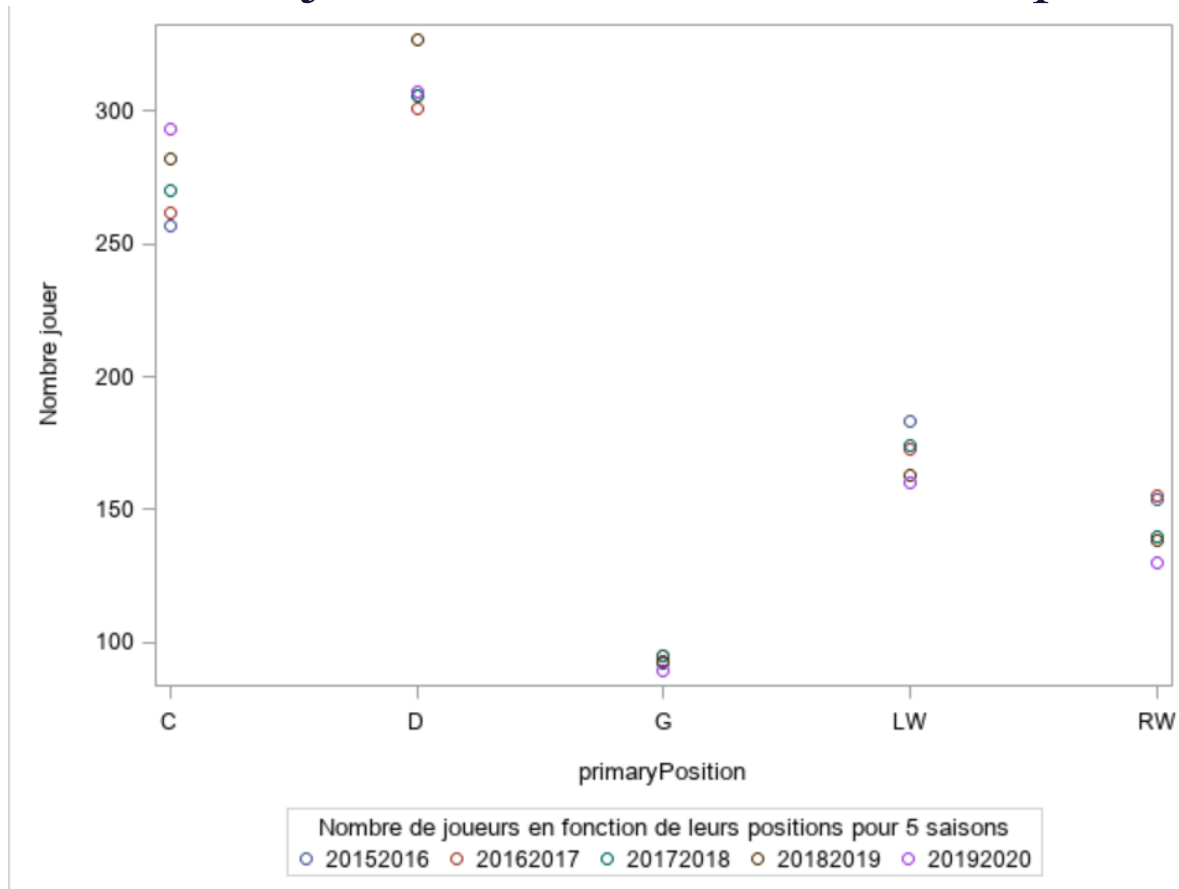
	primaryPosition	avg_timeOnIce	avg_assists	avg_goals	avg_shots
1	C	828.45884131	0.2325796552	0.1599850926	1.5256579382
2	D	1065.8146041	0.2032296503	0.0595392986	1.3068146756
3	G	3154.7540979	0.0161500807	0.0000598659	26.858989861
4	LW	800.47657637	0.2092371875	0.1604003804	1.605731478
5	RW	799.95079883	0.2108213839	0.1712753884	1.6505620074

```
proc sql;  
create table Projet.nbe_postes as  
select season,primaryPosition,count(*) as nbe_player  
from Projet.ind_moy_20152020  
group by season,primaryPosition  
;
```

```
create table Projet.pos_moy_20152020 as  
select DISTINCT primaryPosition,  
                avg(avg_timeOnIce) as avg_timeOnIce,  
                avg(avg_assists) as avg_assists,  
                avg(avg_goals) as avg_goals,  
                avg(avg_shots) as avg_shots  
from Projet.ind_moy_20152020  
Group by primaryPosition  
;  
quit;
```


Analyse des données : Etude de graphiques

1. Nombre de joueurs en fonction de leurs positions pour 5 saisons



Pour commencer l'analyse des données, on a voulu tout d'abord observer les proportions des joueurs dans chaque position par saison.

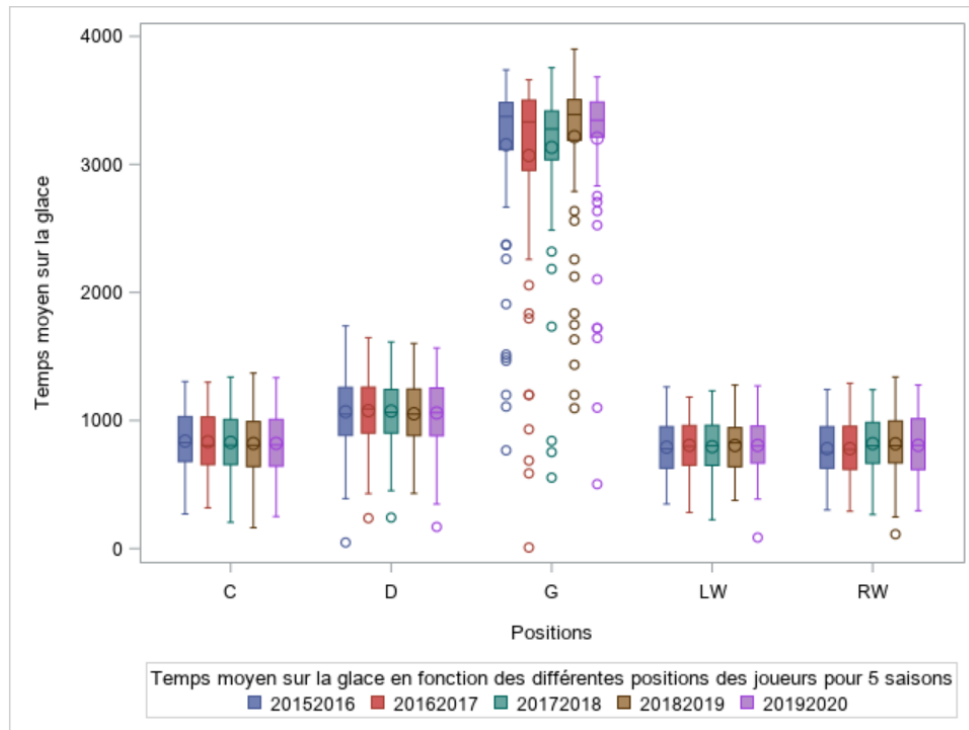
On remarque pour chaque saison, le nombre de joueur de chaque position est différent.

Par exemple, entre 250 et 300 joueurs par saison pour les joueurs au centre ou encore entre 150 et 200 joueurs pour les ailiers droits.

Il y a des disparités quant au nombre de joueurs recrutés sur les différentes positions, mais cela est dû au fait qu'il y ait plus de changement sur certaines positions lors des matchs, ou encore plus de blessés sur certaines actions, etc.

Analyse des données : Etude de graphiques

2. Temps moyen sur la glace en fonction des différentes positions des joueurs pour 5 saisons



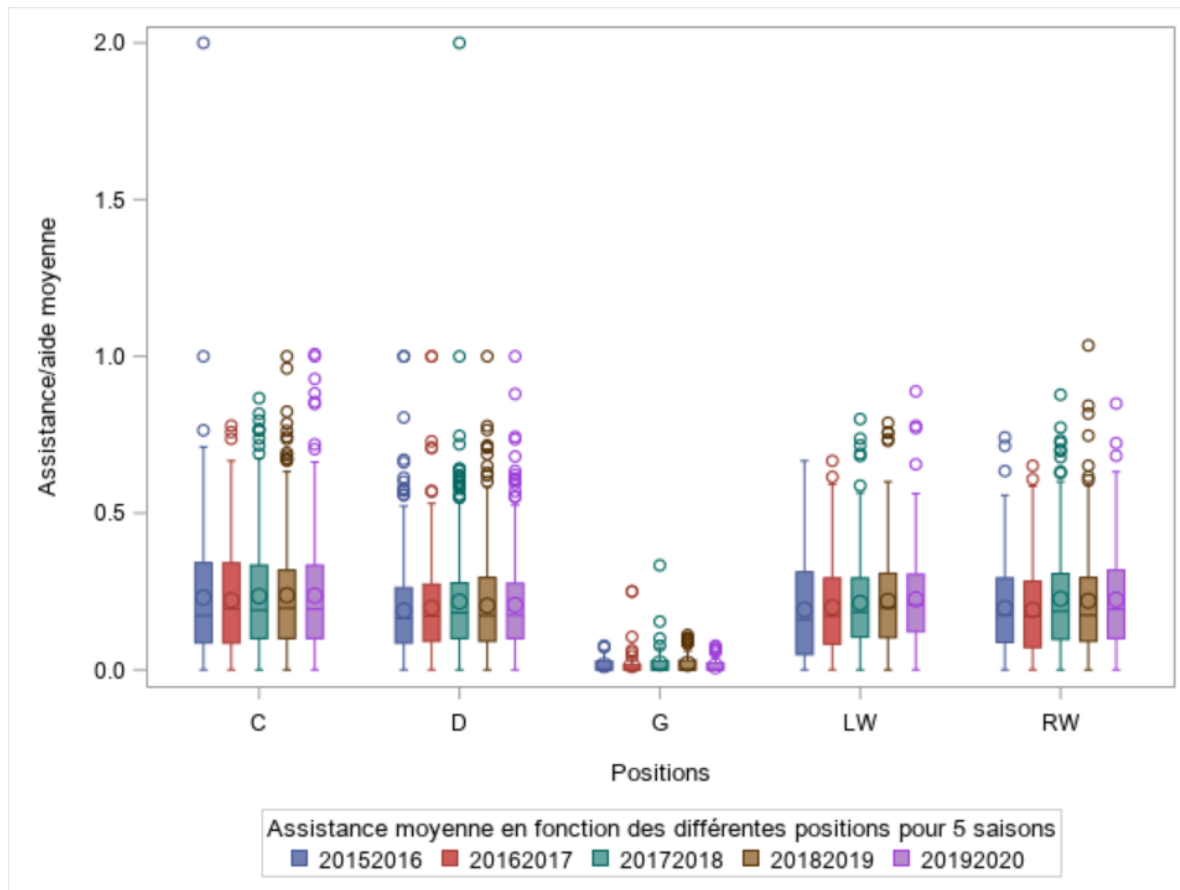
On remarque que les gardiens passent largement plus de temps sur la glace que les autres positions de hockey sur glace, en moyenne entre 3000 et 4000 secondes par match, c'est-à-dire entre 50 et 66 minutes. Cependant, il y a beaucoup plus de dispersion au niveau du temps des différents gardiens passés sur la glace, on le remarque avec les nombreuses valeurs atypiques. De plus, pour la saison 2016-2017, il existe une plus forte dispersion en moyenne des valeurs que les autres années pour les gardiens.

Pour les autres positions, on remarque que pour chaque saison, la dispersion au sein des mêmes positions reste à peu près équivalente, c'est-à-dire que le nombre de secondes passées sur la glace pour chaque poste reste à peu près inchangée et entre 800 et 1200 secondes en moyenne, soit entre 13 minutes et 20 minutes par match.

Cela est dû au fait qu'au hockey sur glace, un match standard est composé de 4 périodes de 15 minutes entre coupées d'une pause de 15 minutes à la fin de la 2^{ème} période, et 2 minutes entre la 1^{ère} et 3^{ème} période. Et comme il y a des changements de joueurs, alors cela explique les différences de temps passées sur la glace.

Analyse des données : Etude de graphiques

3. Assistance moyenne en fonction des différentes positions pour 5 saisons

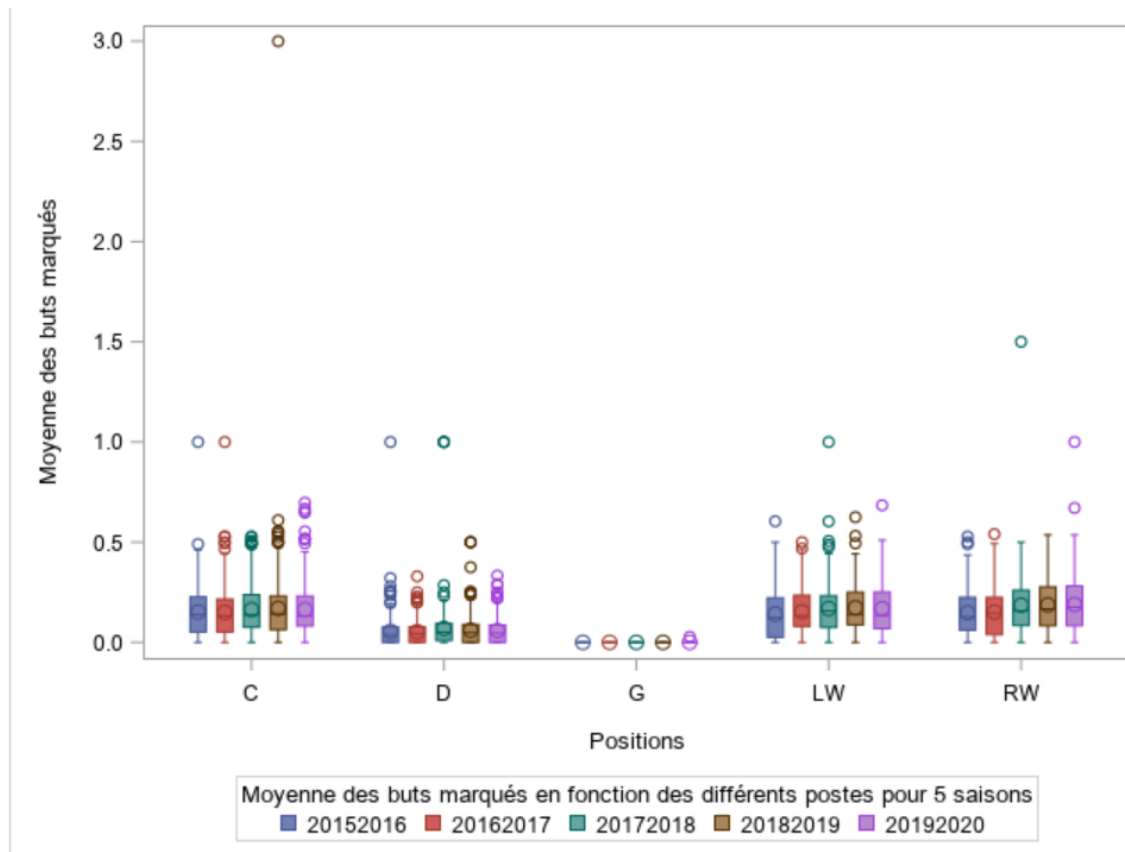


Au hockey sur glace, lorsqu'une équipe inscrit un but, une aide est attribuée aux deux joueurs précédents qui ont touché le palet vers le buteur, la variable "assists" traduit cela dans les données.

On remarque que chacune des positions, hormis les gardiens, aide en moyenne le 'buteur' à marquer de manière équivalente dans les deux dernières passes. Cependant, il existe de nombreuses valeurs atypiques, notamment pour les positions du centre et défenseur à 2.0 d'assistance moyenne pour la saison 2015-2016 et la saison 2017-2018 respectivement.

Analyse des données : Etude de graphiques

4. Moyenne des buts marqués en fonction des différents postes pour 5 saisons



On remarque que les joueurs du centre, les ailiers gauches ainsi que les ailiers droits ont plus tendance à marquer qu'importe les saisons, et cela semble logique, c'est le rôle qui est directement attribué à leur position.

Pour l'année 2018-2019, il y a un individu atypique pour un joueur ayant un poste au centre ayant effectué 3 buts en moyenne, ce qui est largement supérieur à toutes les autres valeurs.

Analyse des données : ACP

1. Analyse en moyenne des variables en fonction de tous les types de joueurs

Matrice de corrélation										
	avg_assists	avg_goals	avg_shots	avg_powerPlayAssists	avg_faceOffWins	avg_faceoffTaken	avg_shortHandedGoals	avg_shortHandedAssists	avg_plusMinus	avg_evenTimeOnIce
avg_assists	1.0000	0.5444	0.6386	0.7831	0.2964	0.2976	0.1580	0.1646	0.3142	0.5013
avg_goals	0.5444	1.0000	0.7335	0.4782	0.3013	0.3057	0.2750	0.0525	0.2544	-0.0414
avg_shots	0.6386	0.7335	1.0000	0.5860	0.2378	0.2387	0.2247	0.0720	0.1592	0.0658
avg_powerPlayAssists	0.7831	0.4782	0.5860	1.0000	0.2344	0.2338	0.0932	0.0744	0.1022	0.0752
avg_faceOffWins	0.2964	0.3013	0.2378	0.2344	1.0000	0.9931	0.2237	0.1006	0.0364	0.1291
avg_faceoffTaken	0.2976	0.3057	0.2387	0.2338	0.9931	1.0000	0.2261	0.0981	0.0341	0.1193
avg_shortHandedGoals	0.1580	0.2750	0.2247	0.0932	0.2237	0.2261	1.0000	0.1429	0.0858	0.2725
avg_shortHandedAssists	0.1646	0.0525	0.0720	0.0744	0.1006	0.0981	0.1429	1.0000	0.1040	0.2821
avg_plusMinus	0.3142	0.2544	0.1592	0.1022	0.0364	0.0341	0.0858	0.1040	1.0000	0.1013
avg_evenTimeOnIce	0.5013	0.2517	0.4737	0.4038	0.0124	0.0048	0.0632	0.1323	0.1558	0.0048
avg_shortHandedTimeOnIce	0.1255	-0.0414	0.0658	0.0752	0.1291	0.1193	0.2725	0.2821	0.1013	0.2237
avg_powerPlayTimeOnIce	0.7221	0.5883	0.6979	0.8197	0.2230	0.2237	0.1097	0.0701	0.0947	0.2237

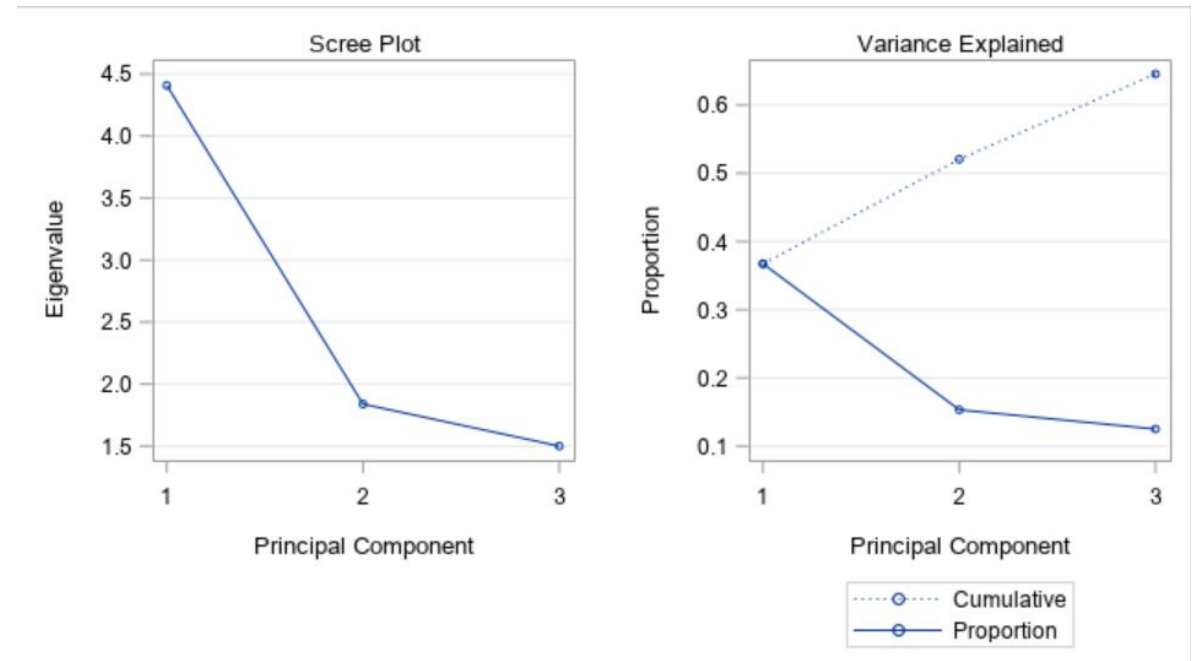
En étudiant la matrice de corrélation de variables de l'ACP, on peut remarquer les liens entre les différentes variables. Par exemple, on remarque que la moyenne des tirs marqués (*avg_shots*) est très corrélée avec la moyenne des buts marqués (*avg_goals*), et cela semble tout à fait logique car sans tirs, il n'y a pas de buts. De plus, la moyenne du temps où l'équipe a plus de joueurs sur la glace que l'équipe adverse (*avg_powerPlayTimeOnIce*) est très corrélée aux moments où l'équipe a un avantage dans l'assistance portée au buteur (*avg_powerPlayAssist*), encore une fois, c'est tout à fait normal que lorsqu'il y ait par exemple un joueur de plus dans l'équipe (car pénalité dans l'équipe adverse) que l'équipe adverse, alors il y ait plus de temps sur la glace pour l'équipe en position de force, et donc elle a plus de chances d'essayer de marquer. Ou encore, il n'y a pas de lien, ou très peu, entre le temps passé sur la glace équivalent pour les joueurs (*avg_evenTimeOnIce*) et la mise en jeu de palet (*avg_faceOffWins* ou *avg_faceOffTaken*)

Analyse des données : ACP

1. Analyse en moyenne des variables en fonction de tous les types de joueurs

La procédure de l'ACP permet d'étudier de manière précise les relations entre des variables quantitatives.

On remarque que sur les deux premiers axes de notre ACP, il y a plus de 52.5% de l'inertie totale, donc il serait aussi intéressant d'analyser le troisième axe.



Analyse des données : ACP

1. Analyse en moyenne des variables en fonction de tous les types de joueurs

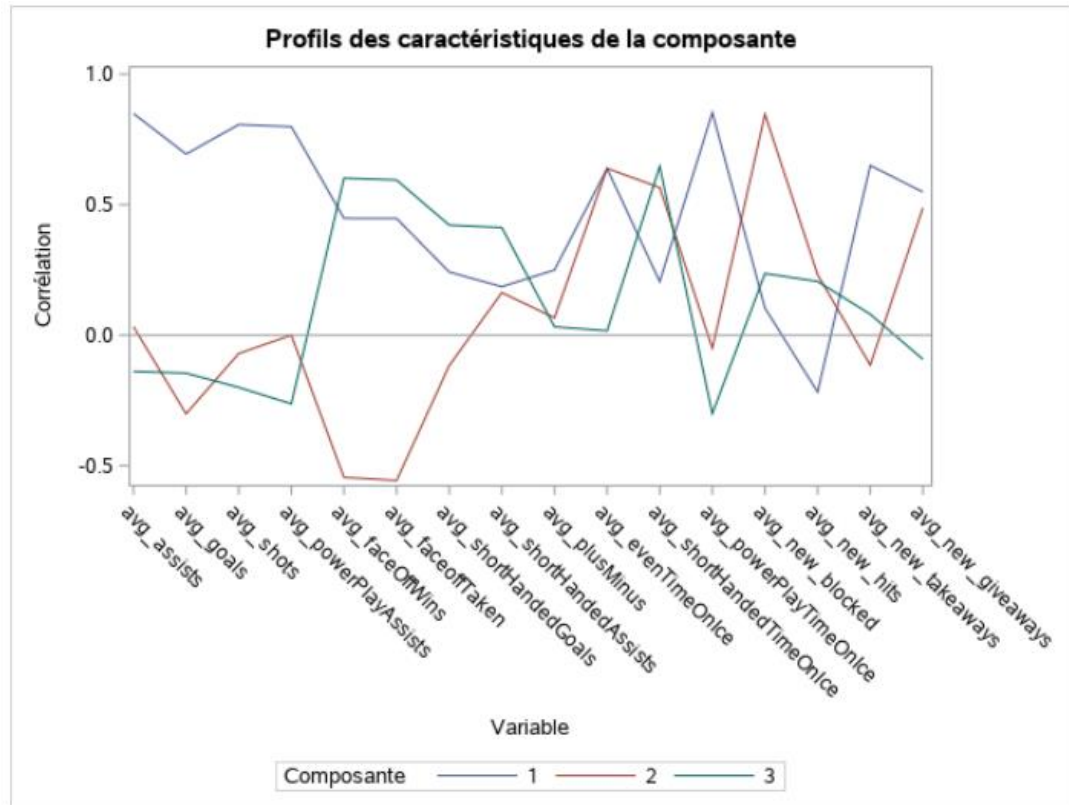
Le premier axe de l'ACP est mesuré majoritairement avec les variables *avg_assists*, *avg_powerPlayTimeOnIce*, *avg_shots*, *avg_powerPlayAssists*, *avg_goals*, donc par les caractéristiques de joueurs de jeu en général.

Alors que le second axe, est mesuré plus en fonction des variables *avg_faceOffWins* et *avg_faceoffTaken* avec un fort taux négatif de *avg_evenTimeOnIce*, c'est-à-dire en fonction de la mise en jeu du palet après avoir gagné ou perdu le point.

Eigenvectors			
	Prin1	Prin2	Prin3
avg_assists	0.410127	-.137806	-.006131
avg_goals	0.356398	-.026092	-.180703
avg_shots	0.392750	-.136092	-.089599
avg_powerPlayAssists	0.384953	-.182038	-.114862
avg_faceOffWins	0.236735	0.600428	-.139046
avg_faceoffTaken	0.236324	0.600878	-.146997
avg_shortHandedGoals	0.140116	0.252742	0.251984
avg_shortHandedAssists	0.104499	0.136926	0.477051
avg_plusMinus	0.132830	-.070327	0.154039
avg_evenTimeOnIce	0.264727	-.255104	0.365302
avg_shortHandedTimeOnIce	0.100302	0.122581	0.662119
avg_powerPlayTimeOnIce	0.402624	-.198555	-.136003

Analyse des données : ACP

2. Analyse de la corrélation en fonction des moyennes des variables pour les composantes



Après avoir expliqué la corrélation de certaines caractéristiques moyennes des joueurs en général dans la partie 1 de l'ACP, regardons à présent la corrélation de celles-ci à travers 3 différentes composantes via le graphique.

On remarque une forte corrélation entre la première composante de l'ACP et les variables *avg_assists*, *avg_goals*, *avg_shots*, *avg_powerPlayTimeOnIce*, *avg_powerPlayAssist*, donc avec les actions commises par les joueurs en général et aussi lorsqu'ils sont en position avantagée par rapport à l'équipe adverse en moyenne.

La deuxième composante est corrélée négativement avec les variables *avg_faceOffWins* et *avg_faceOffTaken*, donc avec la mise en jeu du palet et positivement avec *avg_new_hits*, *avg_new_giveaways* donc avec la prise en possession du palet et des tirs lancés en moyenne.

La troisième composante représente environ l'opposé de la première composante, cela permet de comparer de manière plus simple les différentes caractéristiques associées.

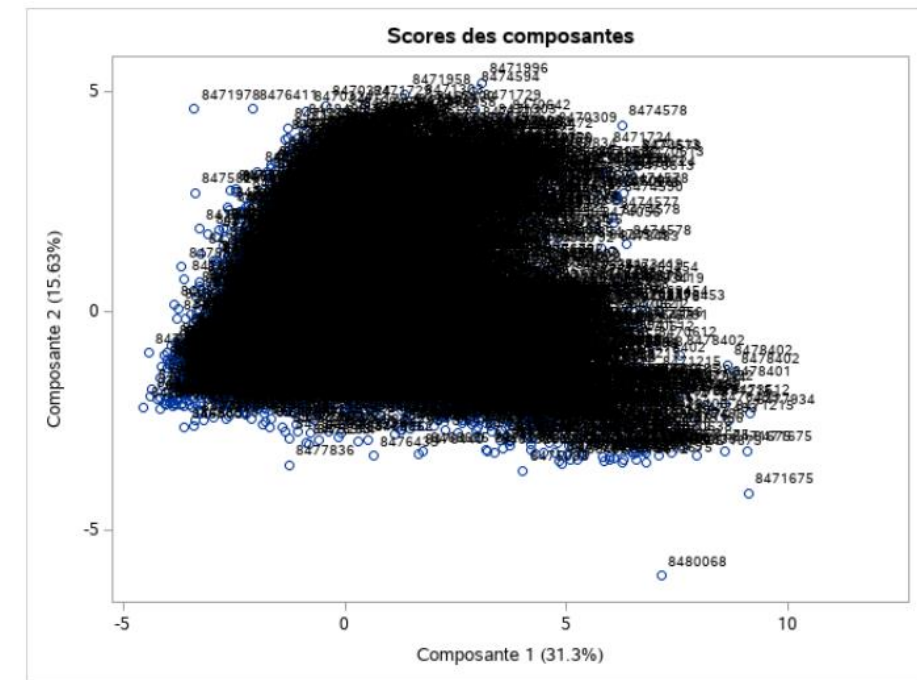
Analyse des données : ACP

3. Analyse de la corrélation en fonction des moyennes des individus sur les composantes

On ne peut pas analyser les individus, il y en a trop.

Cependant, on remarque qu'ils ne se séparent pas en sous-groupe, c'est-à-dire que malgré les positions des joueurs attirées, on ne remarque pas de caractéristiques propres associées dans cette ACP.

On remarque quelques individus atypiques, cependant, nous avons décidé de les garder dans notre ACP car nous trouvions très intéressant de garder tous les profils de joueurs et de ne pas en exclure certains.



Analyse des données : ACP +K-means

1. Catégoriser les joueurs par K Means Clustering

- On utilise le K-means Clustering pour regrouper un ensemble particulier de joueurs en fonction de leurs caractéristiques, en les agrégeant en fonction de leurs similitudes.

À partir la bases des données qu'on a créé, on estime le nombre de clusters à 3.

Les cluster sont regroupés sur la base de la distance maximale entre la graine et les observations.

La valeur R au carré du modèle est de 0,57555 ($> 0,50$). Par conséquent, c'est un modèle moyen.

The FASTCLUS Procedure
Replace=FULL Radius=0 Maxclusters=3 Maxiter=1

Initial Seeds				
Cluster	avg_timeOnce	avg_assists	avg_goals	avg_shots
1	1.000000000	0.000000000	0.000000000	1.000000000
2	0.173175745	0.000000000	1.000000000	0.061224490
3	0.241264132	1.000000000	0.000000000	0.000000000

Criterion Based on Final Seeds = 0.0509

Cluster Summary						
Cluster	Frequency	RMS Std Deviation	Maximum Distance from Seed to Observation	Radius Exceeded	Nearest Cluster	Distance Between Cluster Centroids
1	434	0.0552	0.4672		3	0.8021
2	1677	0.0334	0.9636		3	0.1571
3	2834	0.0547	0.8760		2	0.1571

Statistics for Variables				
Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
avg_timeOnce	0.18657	0.05271	0.920212	11.533193
avg_assists	0.08937	0.06493	0.472359	0.895227
avg_goals	0.04427	0.04024	0.174110	0.210815
avg_shots	0.15651	0.02875	0.966262	28.640020
OVER-ALL	0.13158	0.04858	0.863732	6.338485

Pseudo F Statistic = 15662.40

Approximate Expected Over-All R-Squared = 0.57555

Cubic Clustering Criterion = 154.711

WARNING: The two values above are invalid for correlated variables.

Cluster Means				
Cluster	avg_timeOnce	avg_assists	avg_goals	avg_shots
1	0.8423831121	0.0086332229	0.0000213347	0.5742141429
2	0.1691867210	0.0319425132	0.0229024372	0.0205775602
3	0.2658489876	0.1506947746	0.0534851538	0.0375627952

Cluster Standard Deviations				
Cluster	avg_timeOnce	avg_assists	avg_goals	avg_shots
1	0.0784386953	0.0161543386	0.0004444587	0.0759998259
2	0.0411114613	0.0304671191	0.0409698432	0.0129901761
3	0.0539108324	0.0822543718	0.0427960371	0.0214393585

Analyse des données : ACP + K-means

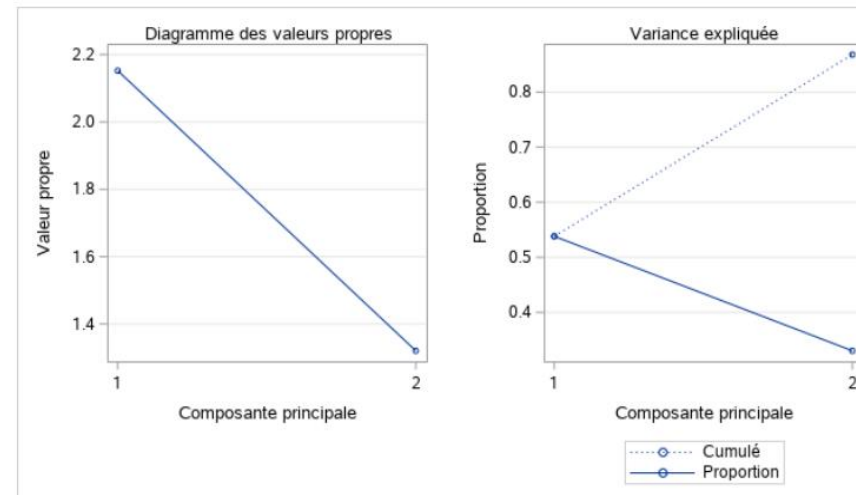
2. Analyse en composantes principales des indicateurs moyens de goalie et skater (Pour toutes les saison entre 2015-2016 et 2019-2020)

Statistiques simples				
	avg_timeOnIce	avg_assists	avg_goals	avg_shots
Moyenne	0.2836676305	0.0979541769	0.0384213493	0.0789020158
StD	0.1865673429	0.0893718120	0.0442687554	0.1565123676

Matrice de corrélation				
	avg_timeOnIce	avg_assists	avg_goals	avg_shots
avg_timeOnIce	1.0000	-.0984	-.1556	0.9483
avg_assists	-.0984	1.0000	0.5164	-.2561
avg_goals	-.1556	0.5164	1.0000	-.2074
avg_shots	0.9483	-.2561	-.2074	1.0000

Valeurs propres de la matrice de corrélation				
	Valeur propre	Différence	Proportion	Cumulé
1	2.15234563	0.83215349	0.5381	0.5381
2	1.32019214		0.3300	0.8681

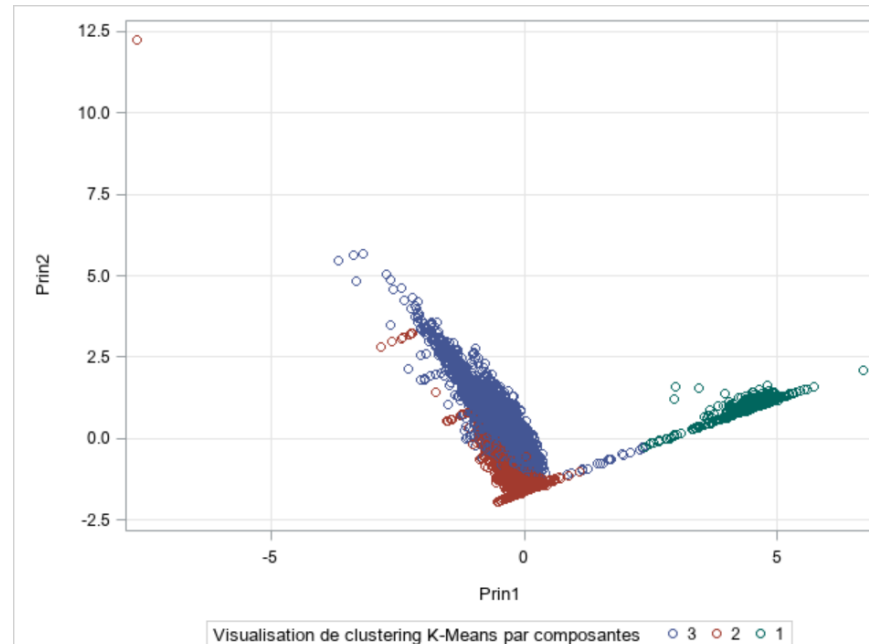
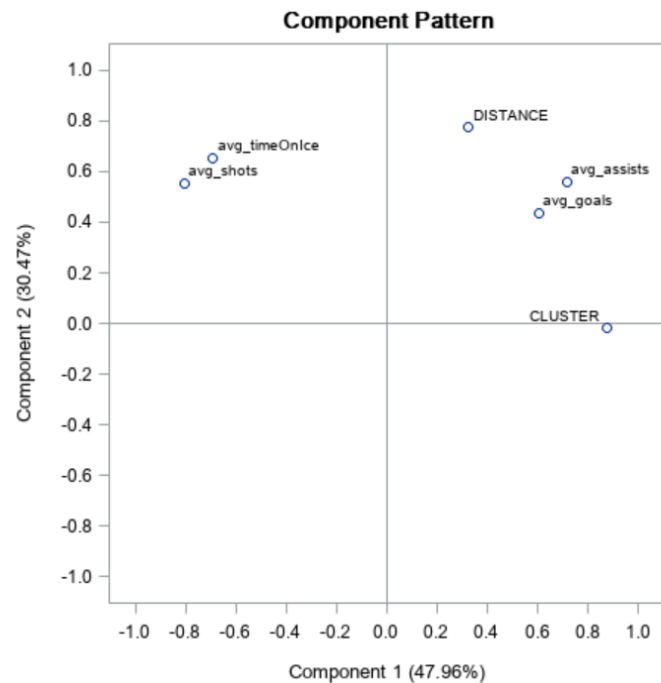
Vecteurs propres		
	Prin1	Prin2
avg_timeOnIce	0.597061	0.400521
avg_assists	-.348466	0.617423
avg_goals	-.350508	0.607429
avg_shots	0.631849	0.299002



Après avoir analysé de façon général toutes les moyennes des caractéristiques des joueurs, il est nécessaire d'en sélectionner certaines afin de pouvoir remarquer les différentes relations entre les positions. Nous avons choisi : *avg_timeOnIce*, *avg_assists*, *avg_goals*, *avg_shots*. On ne regardera que les deux premiers axes de l'ACP car il y a environ 87% de l'inertie.

Analyse des données : ACP + K-means

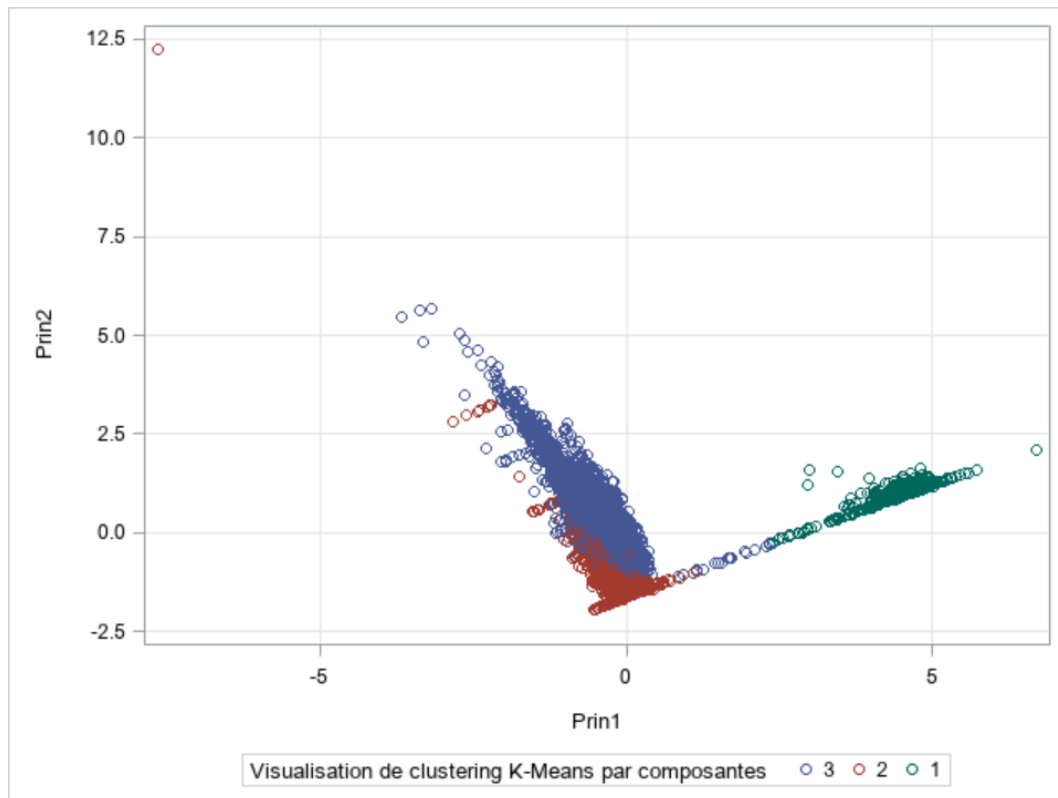
3. Visualiser les clusters par composantes



L'objectif de notre analyse est de déterminer la position de chaque individu via les 4 variables: *avg_timeOnIce*, *avg_assists*, *avg_goals* et *avg_shots*. Nous remarquons qu'il y a donc trois groupes/clusters différents via les 3 couleurs sur le graphique à droite.

Chaque groupe n'est pas associé directement à une position des joueurs en particulier, mais, à partir des analyses faites précédemment, et du graphique à gauche, nous pouvons dire à quel groupe/cluster appartiennent les joueurs et associer cela avec les différentes positions des hockeyeurs.

Analyse des données : ACP + K-means



Il y a donc trois groupes, analysons-les en détails.

- **groupe 1** : joueurs qui marquent des buts, qui portent plus assistance -> on en déduit que ce sont les ailiers droits et gauches (car ils ont à peu près les mêmes compétences) qui sont regroupés : ils doivent aider l'équipe à marquer ainsi qu'essayer de marquer eux-mêmes.
- **groupe 2** : joueurs qui passent du temps sur la glace, marquent, portent assistance et tirs -> on peut en déduire que ce sont les joueurs au centre car ils sont très polyvalents et doivent exceller dans toutes les compétences que doivent avoir les hockeyeurs.
- **groupe 3** : joueurs qui passent beaucoup plus de temps sur la glace que les autres joueurs et tirs aussi beaucoup le palet -> on peut en déduire que ce sont les gardiens et les défenseurs qui sont regroupés car ils doivent défendre leur équipe en dégageant le palet le plus possible, et les gardiens sont sur la glace en moyenne sur toute la durée du match.

Cette représentation permet très clairement de voir les caractéristiques associées à chacune des positions pour les 4 variables *avg_timeOnIce*, *avg_assists*, *avg_goals* et *avg_shots*.

CONCLUSION

Le but de cette analyse est de comprendre les rôles dans l'équipe de Hockey en fonction des résultats réels.

Une équipe a besoin d'une répartition claire des postes, nous réalisons que leur contribution au jeu est tout aussi importante, mais ils assument toujours des responsabilités différentes, détaillées comme suit :

Alors que les gardiens et les défenseurs ont été les plus grands contributeurs au temps de jeu, ce sont les ailiers gauches, les ailiers droits, et les joueurs au centre qui ont marqué le plus. D'autre part, à travers tous les aspects d'un match, les apports de leurs rôles n'ont pas de différence notable (nuage de points ACP).

Les résultats de la classification montrent que dans des aspects tels que le temps moyen passé sur la glace, le nombre moyen de tirs lancés, la moyenne d'assistance portée, le nombre moyen de buts mis, le groupe des ailiers est clairement séparé. Assister, tirer et marquer les buts sont des compétences essentielles pour les joueurs au centre, mais pour les ailiers gauches et les ailiers droits, ces compétences sont plus exigeantes.