

GOLF

Analyse de données exploratoire

Mohamed El Boukhsaibi et Paul Vidal

2024-01-15

Table des matières

Introduction - Présentation de la base de données	3
Variables quantitatives	3
Variables qualitatives	4
Etude descriptives des quantitatives	4
Etudes descriptives des joueurs	5
ACP	6
Matrice des corrélations	6
Valeurs propres	6
Inertie aux axes	7
Etude de variables - Axe 1 et 2	8
Etude de variables - Axe 1 et 3	9
Etude de variables - Axe 2 et 4	10
Etude de variables - Axe 1 et 5	11
Etude des corrélations	12
Nuages de individus	14
Nuages des variables	14
AFC - Fréquentation et mois	15
Tableau de contingence	16
Test du Chi 2	16
Contribution du Chi 2	16
Inertie aux axes	17
Diagramme	17
Table	17
Profil Colonne - Fréquentation	18
Tableau de proportion	18
Diagramme	18
Contribution aux axes	19
Profil ligne - Mois	20
Tableau de proportion	20
Diagramme	20
Contribution aux axes	21
Conclusion AFC 1	22
AFC - Fréquentation et Jour de la semaine	23
Tableau de contingence	23
Conclusion AFC 2	23

AFC - Fréquentation et Météo	24
Tableau de contingence	24
Conclusion AFC 3	24
ACM	25
Inertie des axes	25
Diagramme des inerties	26
Nuages des variables - Dim 1, 2	26
Etude des contribution	27
Nuage des contributions des individus	28
Nuages des variables - Dim 3, 4	30
Etudes des contributions	30
Clustering	32
Choix des cluster	32
Dendogramme	32
Visualisation des clusters	33
Parangons	34
Cluster 1	34
Cluster 2	35
Cluster 3	36
Cluster 4	37
Cluster 5	38
Cluster 6	39
Conclusion générales	40

Introduction - Présentation de la base de données

Dans le monde dynamique du golf, comprendre les interactions entre les conditions environnementales, les caractéristiques temporelles et les comportements des joueurs est essentiel pour améliorer l'expérience de jeu et la gestion des terrains de golf. Le travail présenté ici se penche sur un ensemble de données exhaustif et méticuleusement construit par Samy Baladram, qui vise à capturer l'essence de ces interactions dans le cadre du golf.

L'ensemble de données englobe une période de 3 ans, s'étalant du 1er janvier 2021 au 31 décembre 2023, offrant ainsi une vue d'ensemble complète et diversifiée des habitudes de jeu. La base de données se concentre sur sept joueurs de golf distincts, chacun avec des préférences et des styles de jeu qui leur sont propres, reflétant une variété de profils typiques des joueurs de golf. Les données, bien que générées synthétiquement, s'inspirent étroitement des conditions réelles de Sendai, au Japon, où réside l'auteur, et visent à simuler fidèlement les facteurs influençant le jeu de golf.

Ce compte rendu explore les diverses variables quantitatives et qualitatives enregistrées. Les variables quantitatives incluent des mesures telles que la température, le taux d'humidité, le niveau de fréquentation du golf, et les heures de jeu pour chaque joueur. Parallèlement, les variables qualitatives telles que le jour de la semaine, le mois, la saison, les jours fériés, la présence de vent et les conditions météorologiques générales sont également considérées. De plus, le niveau de fréquentation est catégorisé en cinq niveaux, offrant ainsi une granularité plus fine dans l'analyse des tendances de fréquentation.

La problématique centrale de cette étude est de déterminer comment les conditions météorologiques et les caractéristiques temporelles influencent la fréquentation du golf et le temps de jeu des joueurs. En examinant ces relations complexes, l'étude vise à apporter des éclairages précieux sur les préférences et les comportements des joueurs dans différents contextes environnementaux et temporels.

La base de données rassemblent les variables ci-dessous :

Variables quantitatives

Nom	Description
Temperature	La température extérieur
Humidity	La taux d'humidité ambiant
Crowdedness	Le taux de fréquentation du golf
PlayTimeHour_A	Nombre d'heure de jeux de A
PlayTimeHour_B	Nombre d'heure de jeux de B
PlayTimeHour_C	Nombre d'heure de jeux de C
PlayTimeHour_D	Nombre d'heure de jeux de D
PlayTimeHour_E	Nombre d'heure de jeux de E
PlayTimeHour_F	Nombre d'heure de jeux de F
PlayTimeHour_G	Nombre d'heure de jeux de G

Variables qualitatives

Nom	Description
Weekday	Le jour de la semaine (de lundi à vendredi)
Month	Le mois
Season	La saison
Holiday	1 si le jour est férié, 0 sinon
Windy	1 s'il y a du vent, 0 sinon
Outlook	Aperçu (soleil, couvert, pluie, neige)
Crowdedness_cat	Niveau de fréquentation en 5 niveau (très faible à très élevé)
Play_A	1 si le joueur A à joué, 0 sinon
Play_B	1 si le joueur B à joué, 0 sinon
Play_C	1 si le joueur C à joué, 0 sinon
Play_D	1 si le joueur D à joué, 0 sinon
Play_E	1 si le joueur E à joué, 0 sinon
Play_F	1 si le joueur F à joué, 0 sinon
Play_G	1 si le joueur G à joué, 0 sinon

Etude descriptives des quantitatives

	Temp	Humidity	Crowdedness
min	-2.0	18.0	0.0
max	29.5	99.0	1.0
mean	13.4	61.5	0.6
median	13.8	61.0	0.6
1st	6.0	52.0	0.5
3rd	20.5	72.0	0.7

Les statistiques descriptives des données quantitatives montrent que la température fluctue entre des extrêmes de -2.0°C et 29.5°C et l'humidité varie largement de 18% à 99% indiquant une large amplitude thermique susceptible d'affecter le confort de jeu et potentiellement la fréquentation. Le terrain de golf présente un taux de fréquentation moyen de 0.6 sur une échelle de 0 à 1, suggérant une occupation modérée à élevée. En somme, le terrain de golf connaît des conditions climatiques variées et une fréquentation avec une tendance vers une occupation modérée à élevée. Ces conditions pourraient avoir un impact significatif sur le temps de jeu et les préférences des joueurs.

Etudes descriptives des joueurs

	A	B	C	D	E	F	G
Total	229.00	245.00	256.00	208.00	282.00	66.00	113.00
Moyenne	3.34	3.29	4.64	3.37	3.07	3.31	3.36
Ecart type	0.22	0.20	0.18	0.23	0.24	0.23	0.24
Min	3.10	3.10	4.40	3.10	2.80	3.10	3.10
Quartile1	3.20	3.10	4.50	3.20	2.90	3.10	3.20
Mediane	3.30	3.20	4.60	3.30	3.00	3.20	3.30
Quartile3	3.50	3.40	4.80	3.50	3.20	3.40	3.50
Max	4.10	3.90	5.20	4.20	4.00	3.90	4.20

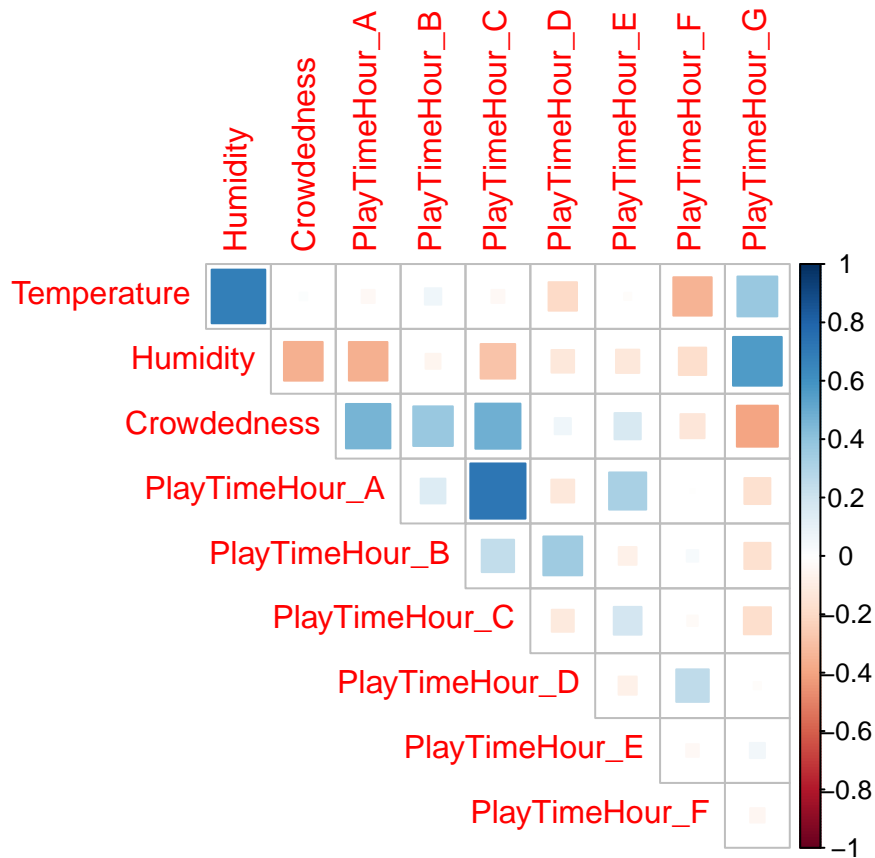
Les moyennes des temps de jeu varient de 3.07 à 4.64 heures, ce qui montre que les joueurs C et E tendent à jouer plus longtemps par session que les autres. Les écarts-types sont tous inférieurs à 0.25, ce qui suggère une cohérence dans les habitudes de jeu de chaque joueur. Le joueur C semble être le plus engagé, avec des sessions de jeu plus longues et un total d'heures plus élevé, tandis que le joueur F a le total d'heures le plus faible, ce qui pourrait indiquer une fréquentation moins régulière. Les quartiles et les médianes suggèrent que la majorité des sessions de jeu pour chaque joueur se concentrent autour d'une durée médiane spécifique, avec peu de variations extrêmes, comme indiqué par les valeurs minimales et maximales proches des médianes. Ces statistiques démontrent les tendances de chaque joueur concernant la durée de leurs sessions de jeu et peuvent être indicatives de leurs préférences ou contraintes personnelles.

ACP

Après avoir établi le cadre de notre étude sur les données de golf et identifié la question centrale de l'influence des conditions météo et temporelles sur le jeu, nous nous apprêtons à approfondir notre analyse grâce à l'Analyse en Composantes Principales (ACP). Cette méthode de réduction de la dimensionnalité va nous permettre de clarifier la structure de nos données quantitatives, en extrayant les facteurs les plus influents sur la fréquentation et le temps de jeu.

Matrice des corrélations

Présentons dans un premier temps la matrice des corrélations :

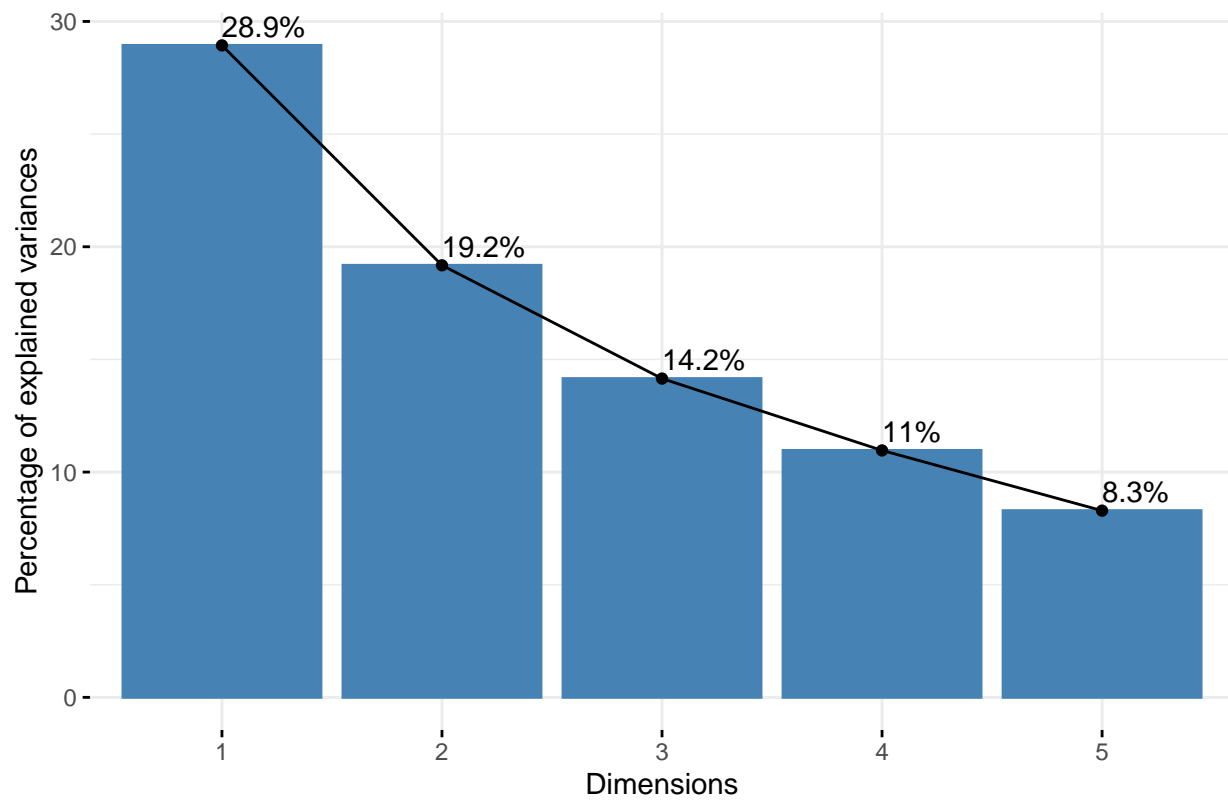


Valeurs propres

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	3	29	29
comp 2	2	19	48
comp 3	1	14	62
comp 4	1	11	73
comp 5	1	8	82
comp 6	1	6	87
comp 7	0	5	92
comp 8	0	4	95
comp 9	0	3	98
comp 10	0	2	100

Inertie aux axes

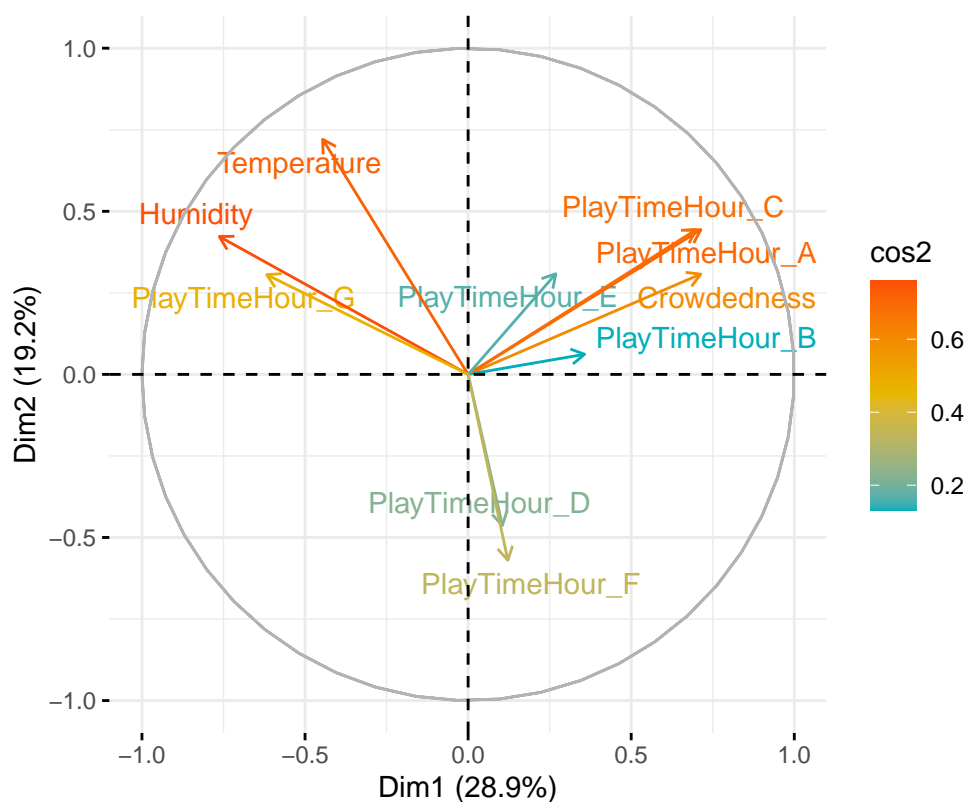
Scree plot



On conserve les 4 premières axes.

Etude de variables - Axe 1 et 2

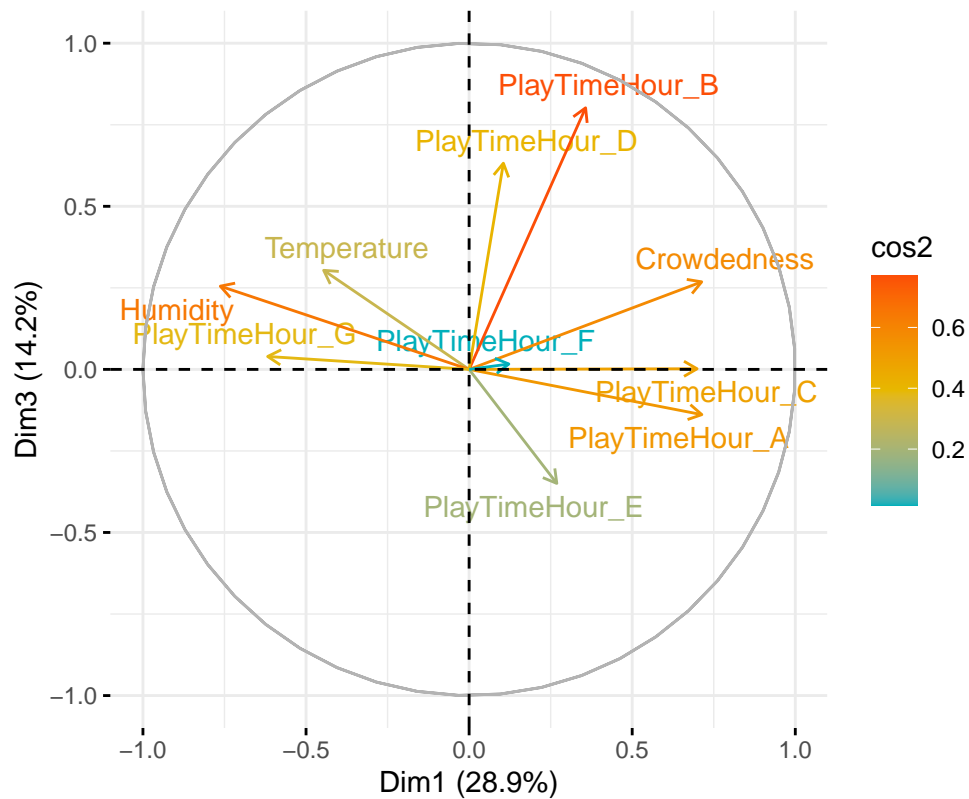
Variables – PCA



Le premier graphique met en relief la corrélation positive entre le temps de jeu des joueurs A et C et la variable “crowdedness” qui met en avant la densité de joueur déjà présents ce jour là. A et C ont donc tendance à jouer au golf les jours où pas mal de monde se trouve déjà présents sur place. A, C et “Crowdedness” sont placés positivement sur l’axe 1 contrairement à la variable “humidity”, ce qui nous laisse croire que l’humidité est un facteur météorologique qui dérange A, C et d’autres joueurs les contraignant donc à reporter leur partie de golf à plus tard les jours humides. Le joueur G est également placé négativement sur l’axe 1 et positivement sur l’axe 2, nous laissant croire que le joueur G a un temps de jeu plus long les jours humides et chauds, et n’est pas très pratiquant de golf les jours où il y a beaucoup de monde sur place. G a donc des exigences opposés à A et C concernant ses paramètres de jeu. Cela peut se traduire de la sorte : le fait que G n’aime pas jouer au golf lorsqu’il y a beaucoup de monde le contraint à jouer les jours humides et chauds. Cela dit, il est tout de même important de préciser que les vecteurs A et C se superposent et sont bien représentés sur ce graphique, nous indiquant la forte corrélation entre le temps de jeu des ces deux individus. Peut-être jouent-ils ensemble ?

Etude de variables - Axe 1 et 3

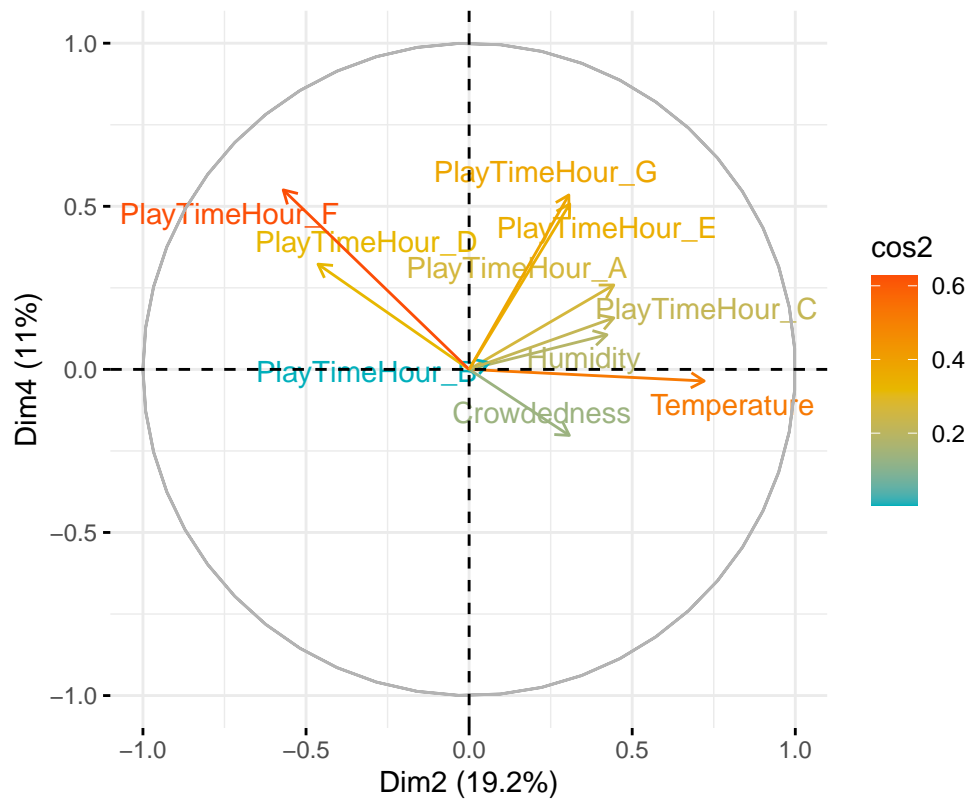
Variables – PCA



Ici le temps de jeu de B est bien représenté, et on voit, de part son placement positif sur la dimension 1 qu'il aime jouer les jours où il y a beaucoup de monde. On remarque aussi graphiquement la corrélation positive entre le temps de jeu de B et celui de D, peut-être aiment-ils jouer ensemble ? On retrouve également l'opposition de la variables crowdedness et humidity, indiquant que les joueurs en générale n'aime pas vraiment jouer au golf les jours humides. Le temps de jeu de B et la variable crowdedness sont ici placés positivement sur l'axe 1, indiquant peut-être que le temps de jeu de B est plus long les jours où il y a beaucoup de monde.

Etude de variables - Axe 2 et 4

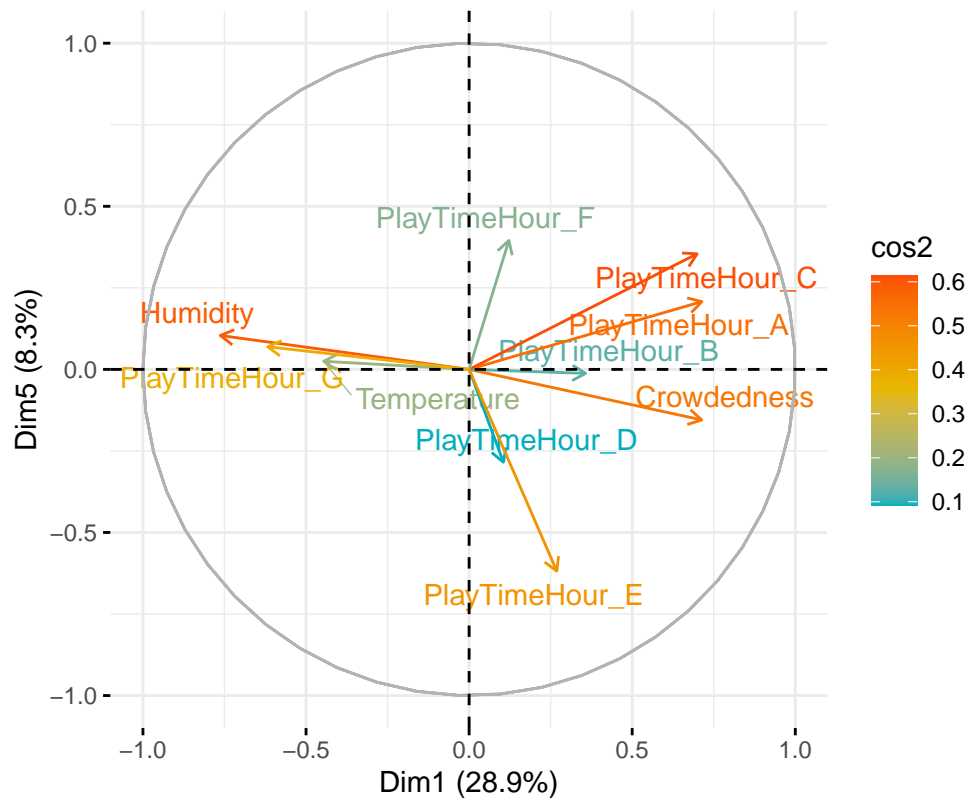
Variables – PCA



Lorsqu'on mets en avant l'axe 2 avec l'axe 4, on remarque assez aisément la corrélation négative entre le temps de jeu de F et la température; la conclusion est alors assez simple, plus il fait chaud, moins F a tendance à jouer longtemps. Plus F joue longtemps, plus D joue longtemps, peut être aiment-t-ils jouer ensemble, ou ont-t-ils des exigences météorologiques similaires concernant leurs parties de golf.

Etude de variables - Axe 1 et 5

Variables – PCA



La contribution et la corrélation de l'individu E à l'axe 5 sont assez prononcés. Voilà pourquoi il serait intéressant de placer cette axe sur un graphique pour étudier le comportement de ce joueur. Ici on remarque que le temps de jeu de E est placé positivement sur l'axe 1, et on rappelle que l'axe 1 est fortement corrélé aux variables "temps de jeu de A" "temps de jeu de B", "crowdedness" et "humidity". Il est alors assez simple de remarquer que E n'aime pas jouer les jours humides, son temps de jeu sera plus long les jours où déjà beaucoup de monde sont présents (parce que peu de personne n'aime l'humidité pour jouer) et que son temps de jeu est corrélé positivement au temps de jeu de A et C.

Etude des corrélation

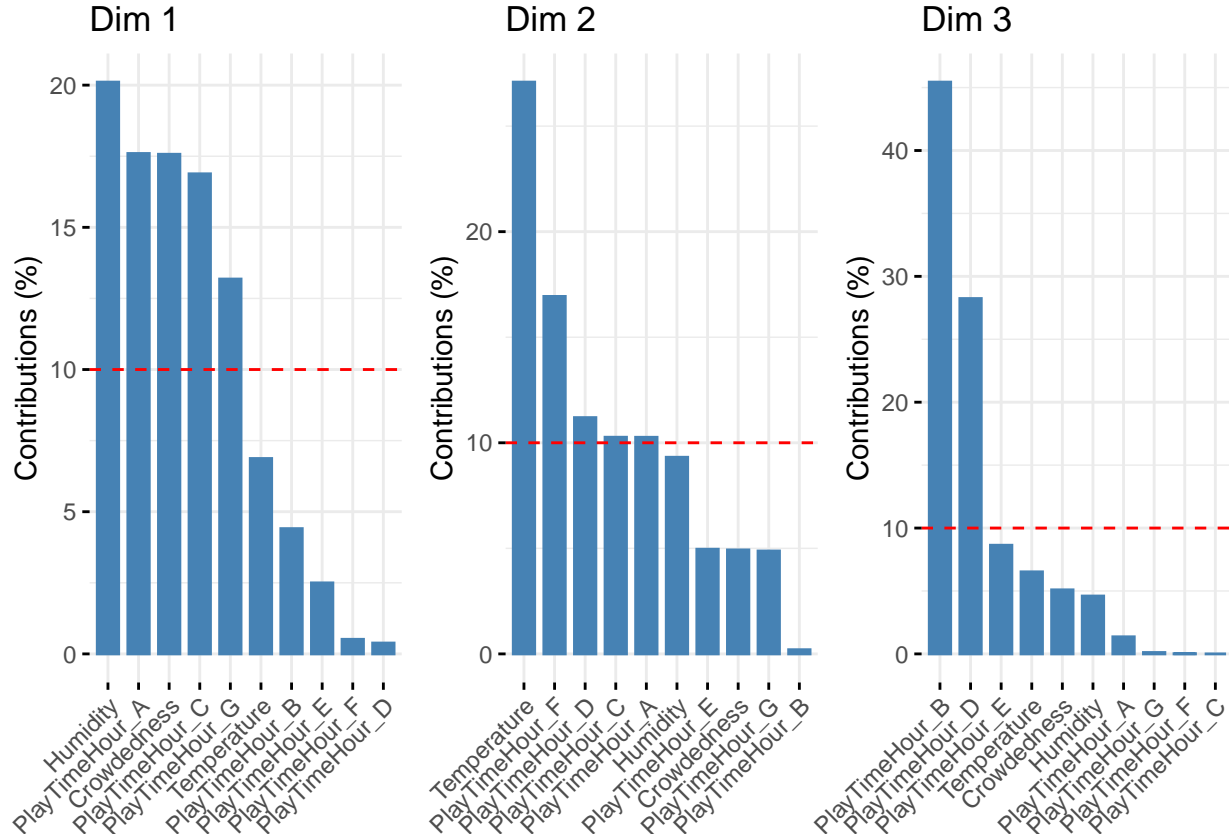
Ici le tableau ci-dessous représente la corrélation entre chaque variable quantitative et chaque dimension.

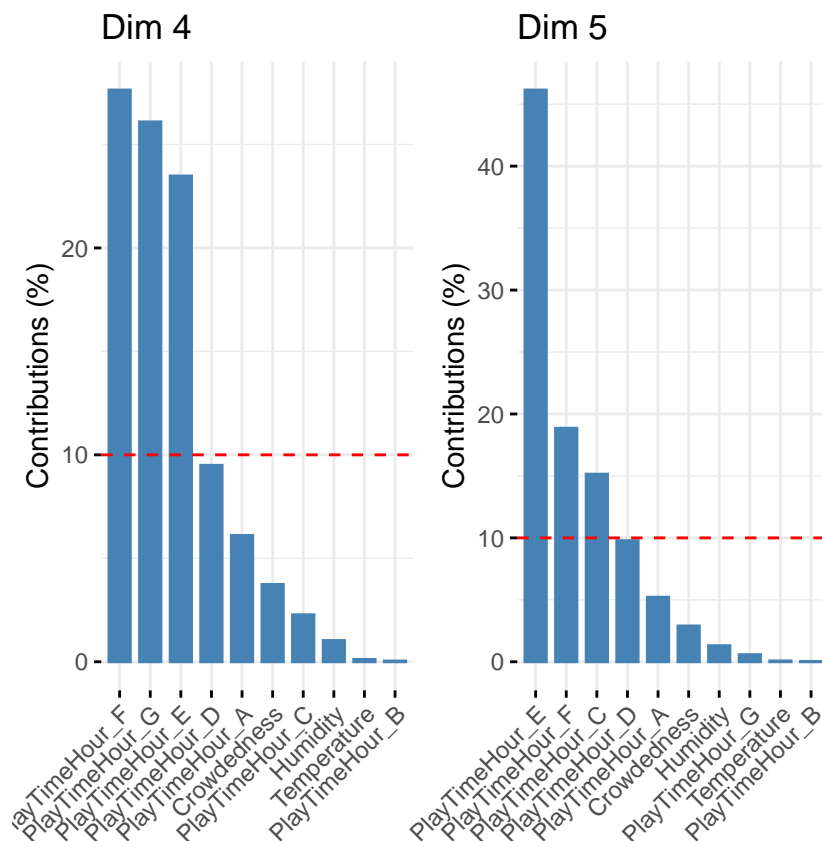
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Temperature	-0.446	0.721	0.304	-0.036	0.025
Humidity	-0.763	0.423	0.255	0.106	0.104
Crowdedness	0.713	0.307	0.268	-0.202	-0.155
PlayTimeHour_A	0.714	0.444	-0.139	0.259	0.208
PlayTimeHour_B	0.357	0.062	0.802	0.021	-0.012
PlayTimeHour_C	0.699	0.444	0.002	0.158	0.354
PlayTimeHour_D	0.105	-0.463	0.632	0.323	-0.285
PlayTimeHour_E	0.269	0.308	-0.350	0.507	-0.618
PlayTimeHour_F	0.122	-0.570	0.016	0.550	0.395
PlayTimeHour_G	-0.618	0.306	0.040	0.535	0.069

Au vu du tableau des corrélations, on observe que :

- L'axe 1 représente bien les variances des variables "humidity", "crowdedness" et du temps de jeu des individus A, C et G
- L'axe 2 explique plus la variance des variables "temperature", "windy" et du temps de jeu de D
- "Weekday" et le temps de jeu de B sont des attributs déterminants de l'axe 3.
- L'axe 4 et 5 représenteront quant à eux beaucoup plus les variances de temps de jeu des joueurs F , G et E.

Affichons les contributions des variables aux dimensions à travers des graphiques afin d'avoir une meilleure visibilité des variables qui contribuent le plus aux différentes axes.





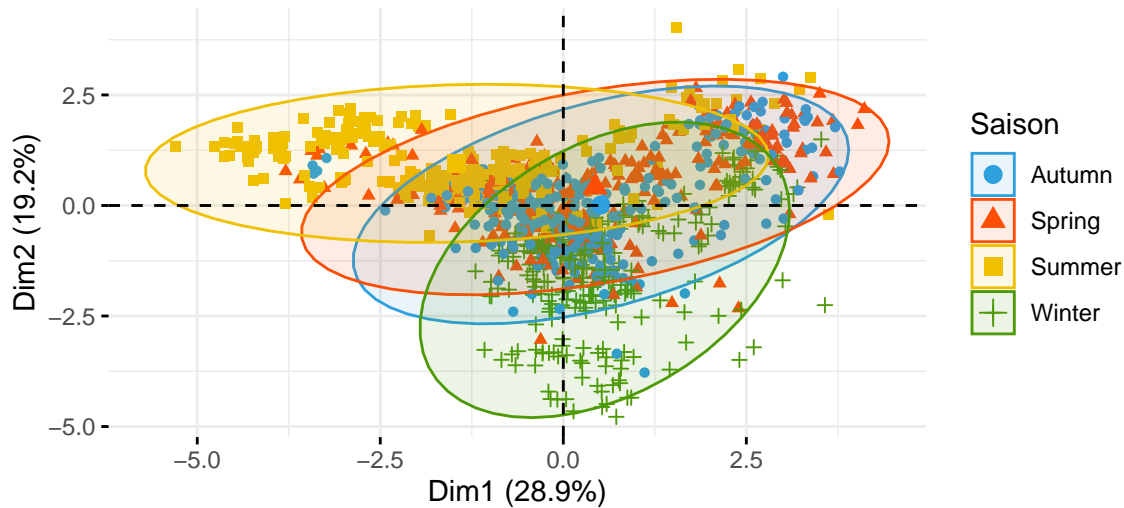
Pour répondre partiellement à la problématique, il est évident que les conditions météorologiques influencent le temps de jeu des différents joueurs présents dans notre base de données. L'ACP nous permet de voir que les joueurs A et C favorisent les jours non humides et ont tendance donc à jouer les jours où il y a pas mal de monde sur place. Leur temps de jeu sont également corrélés entre eux positivement ce qui nous laisserais croire que plus A joue plus C joue et inversement. Peut-être jouent-t-ils ensemble ? Quoiqu'il en soit, leurs conditions de jeu sont similaires. Le temps de jeu du joueur B n'est que peu influencé par les conditions météorologiques, il a cependant un temps de jeu assez long les jours où beaucoup de joueurs sont déjà sur place. Peut-être sont-ils les jours de la semaine qui influencent son temps de jeu ? Les fortes températures et l'humidité constituent des éléments météorologiques qui poussent le joueur D à faire autre chose que jouer une partie de golf. Il est important de noter que le temps de jeu de B et D sont corrélés positivement, indiquant peut-être des préférences météorologiques similaires où des disponibilités dans le temps assez semblables.

Le joueur E, contribuant majoritaire de l'axe 5, a lui des goûts assez similaire en terme de météo que A et C.

F préfère jouer les jours froids, G quant à lui préfère les jours chauds et humides. Il est d'ailleurs le seul joueur de notre base de données à avoir ces préférences. Mais ceci s'explique assez aisément, G n'a pas l'air d'apprécier les terrains de golf très fréquentés pour ses parties. Il s'y rend donc les jours humides et chauds et est sûr de trouver peu de monde, étant donnée que la variable crowdedness et humidity sont corrélés négativement. Tout les deux ont des temps de jeu plus longs lorsqu'il y a moins de monde sur les terrains de golf.

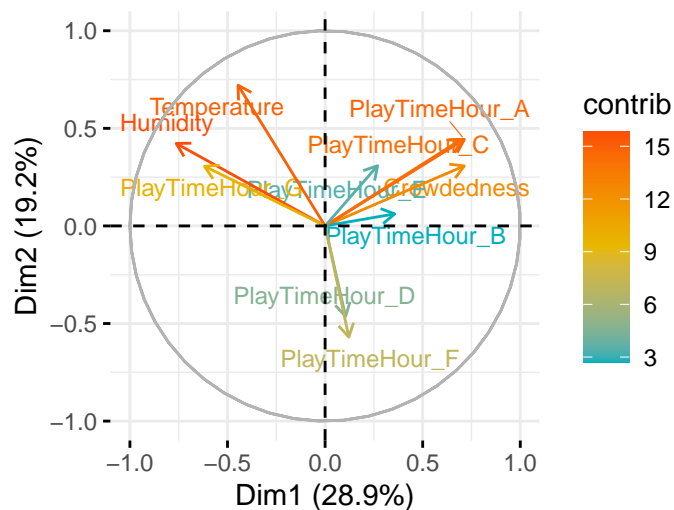
Nuages de individus

Individuals – PCA



Nuages des variables

Variables – PCA



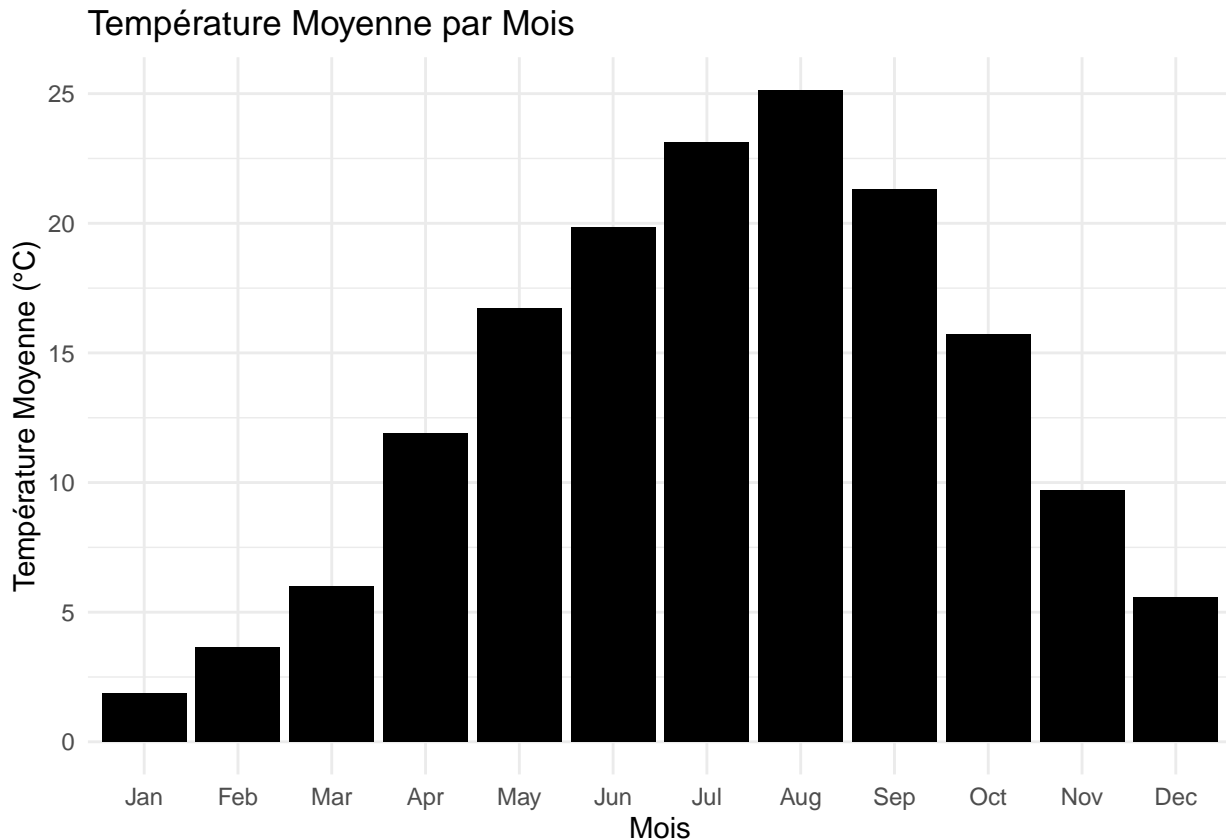
Sans trop de surprise, on remarque que les jours d'été sont placés pour la plupart négativement sur l'axe 1, et rejoins donc les variables température et humidité. À l'inverse, les jours de printemps et d'automne viennent en partie se placer positivement sur la dimension 1 en suivant la direction des temps de jeu de A et C, ainsi que la variable Crowdedness. L'hiver, placé négativement sur l'axe 2, vient suivre la trajectoire des temps de jeu de D et F. Le printemps et l'automne sont les saisons favorites de la plupart des joueurs pour jouer au golf. A et C vont consacrer leurs temps de jeu à ces saisons-ci. G, du fait qu'il n'aime potentiellement pas les terrains de golf bondés, ou du fait de ces préférences peu communes, va favoriser l'été pour sa pratique du golf. D et F quant à eux vont opter pour l'hiver.

Nous allons maintenant poursuivre notre investigation en examinant les associations entre les variables qualitatives et notre variable d'intérêt : la fréquentation du terrain de golf. Pour ce faire, nous utilisons l'Analyse Factorielle des Correspondances (AFC), qui est particulièrement adaptée à l'étude des relations entre deux variables catégorielles.

Trois AFC distinctes vont être menées afin de creuser plus profondément les liens entre la fréquentation et, respectivement, le mois de l'année, le jour de la semaine, et les conditions météorologiques générales (Outlook). Chacune de ces analyses vise à déchiffrer les patterns de fréquentation en fonction de ces catégories temporelles et environnementales, permettant de révéler des tendances potentielles et des préférences comportementales spécifiques.

AFC - Fréquentation et mois

Avant de commencer l'étude et pour nous aider dans nos futurs interprétations, il semble pertinent de regarder la température moyenne de chaque mois. En effet, les températures sont basées sur le climat d'une région japonaises et ne sont peut-être pas sensiblement similaires à nos températures européennes lors d'une période donnée.



Les températures moyennes de chaque mois semblent être similaires à ce qu'on a l'habitude d'observer en France, avec des températures très élevées en été, très faibles en hiver, et plus modérées au printemps et en automne.

Tableau de contingence

	Très faible	Faible	Moyenne	Élevé	Très élevé	Total
Apr	2	17	18	26	27	90
Aug	47	13	18	6	9	93
Dec	13	26	26	17	11	93
Feb	21	22	20	17	4	84
Jan	43	24	13	11	2	93
Jul	41	17	15	10	10	93
Jun	27	16	16	14	17	90
Mar	17	29	16	18	13	93
May	5	8	27	22	31	93
Nov	6	14	20	30	20	90
Oct	7	10	19	25	32	93
Sep	12	15	19	18	26	90
Total	241	211	227	214	202	1095

Test du Chi 2

X-squared	271.59
df	44.00
p.value	0.00

Contribution du Chi 2

	Très faible	Faible	Moyenne	Élevé	Très élevé	Total
Apr	5.90	0.00	0.01	1.48	2.40	9.78
Aug	12.66	0.50	0.03	3.00	1.43	17.62
Dec	1.00	1.34	0.86	0.03	0.81	4.05
Feb	0.13	0.77	0.14	0.01	3.14	4.18
Jan	9.13	0.76	0.75	1.04	4.93	16.62
Jul	7.58	0.02	0.35	1.35	1.10	10.40
Jun	0.96	0.04	0.14	0.27	0.00	1.41
Mar	0.22	2.52	0.21	0.00	0.37	3.32
May	4.30	2.02	1.14	0.30	4.11	11.87
Nov	3.54	0.24	0.04	3.22	0.26	7.30
Oct	3.26	1.29	0.00	0.94	4.73	10.23
Sep	1.13	0.12	0.00	0.00	1.96	3.21
Total	49.83	9.61	3.67	11.65	25.24	100.00

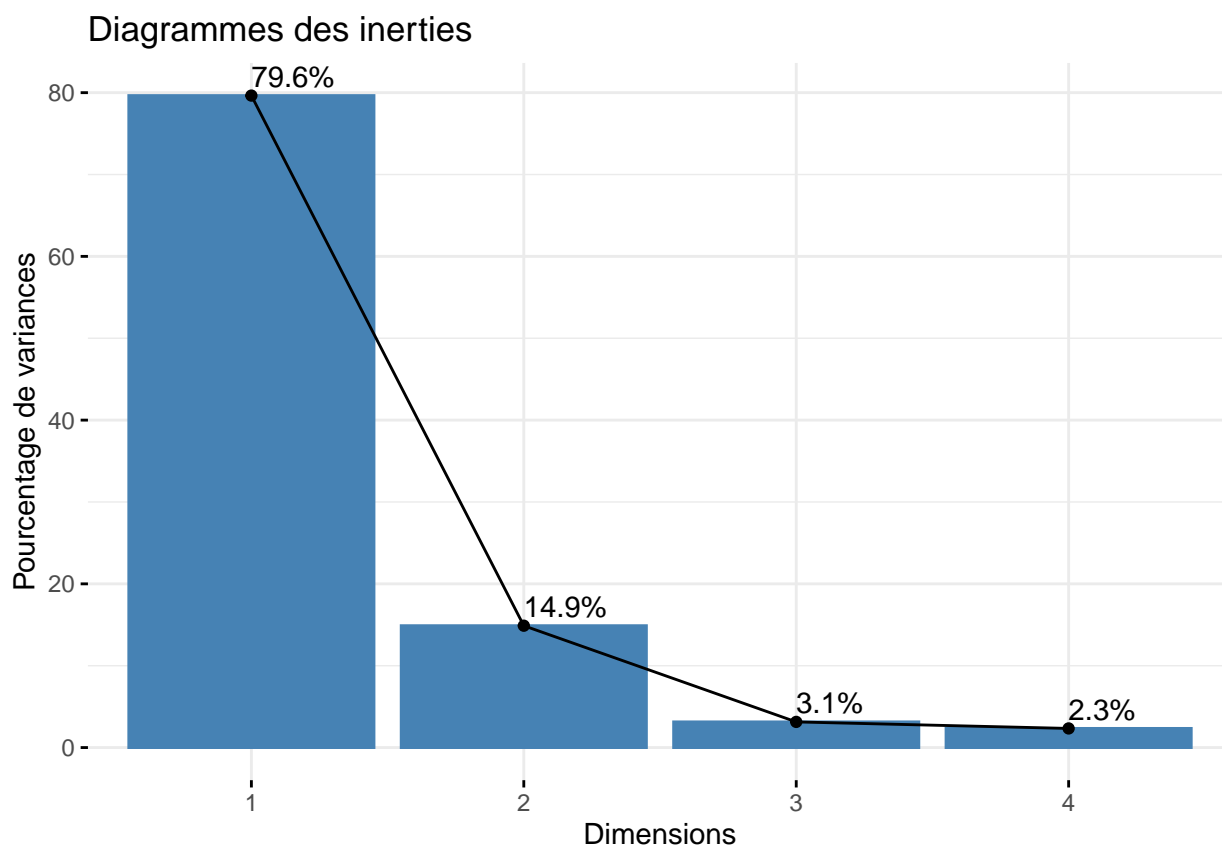
Les résultats du test du Chi-2 pour l'association entre la fréquentation et les mois de l'année révèlent une valeur de Chi-2 de 271.6 avec 44 degrés de liberté et une p-value de 0.0. Cette p-value très faible suggère que l'association entre les mois de l'année et les niveaux de fréquentation est statistiquement significative, ce qui signifie que la fréquentation varie selon le mois.

En examinant les contributions au Chi-2, qui mesurent l'impact de chaque combinaison sur la statistique globale du Chi-2, nous pouvons voir que certains mois et certains niveaux de fréquentation contribuent de manière disproportionnée à la valeur globale du Chi-2. Par exemple, les mois d'août et de janvier présentent des contributions élevées pour les catégories de fréquentation 'Très faible' et 'Très élevé'. Cela peut indiquer que ces mois connaissent des extrêmes en termes de fréquentation, peut-être en raison des conditions saisonnières ou des vacances.

D'autres mois, comme juin et mars, ont des contributions très faibles pour certains niveaux de fréquentation, ce qui pourrait signifier que ces mois n'ont pas une influence aussi forte sur la fréquentation ou que la fréquentation est plus uniformément répartie.

Inertie aux axes

Diagramme



Table

Nous conservons les deux premiers axes qui rassemble 94.5 % de l'inertie totale.

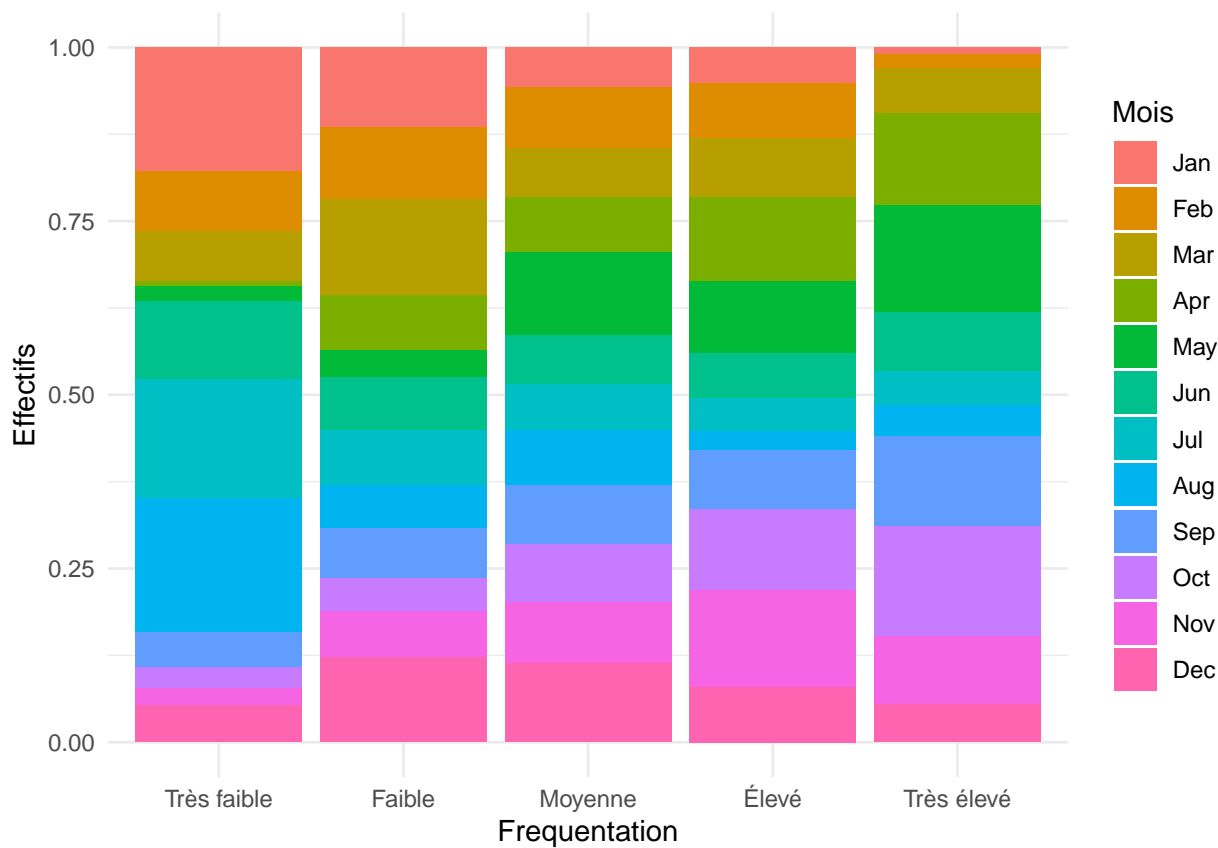
	1	2	3	4
Inerties	0.1975	0.0369	0.0078	0.0058
Inerties relatives %	79.6412	14.8854	3.1359	2.3375
Inerties relatives cumulees	79.6412	94.5266	97.6625	100.0000

Profil Colonne - Fréquentation

Tableau de proportion

	Très faible	Faible	Moyenne	Élevé	Très élevé	Total
Apr	0.8	8.1	7.9	12.1	13.4	8.2
Aug	19.5	6.2	7.9	2.8	4.5	8.5
Dec	5.4	12.3	11.5	7.9	5.4	8.5
Feb	8.7	10.4	8.8	7.9	2.0	7.7
Jan	17.8	11.4	5.7	5.1	1.0	8.5
Jul	17.0	8.1	6.6	4.7	5.0	8.5
Jun	11.2	7.6	7.0	6.5	8.4	8.2
Mar	7.1	13.7	7.0	8.4	6.4	8.5
May	2.1	3.8	11.9	10.3	15.3	8.5
Nov	2.5	6.6	8.8	14.0	9.9	8.2
Oct	2.9	4.7	8.4	11.7	15.8	8.5
Sep	5.0	7.1	8.4	8.4	12.9	8.2
Total	100.0	100.0	100.0	100.0	100.0	100.0

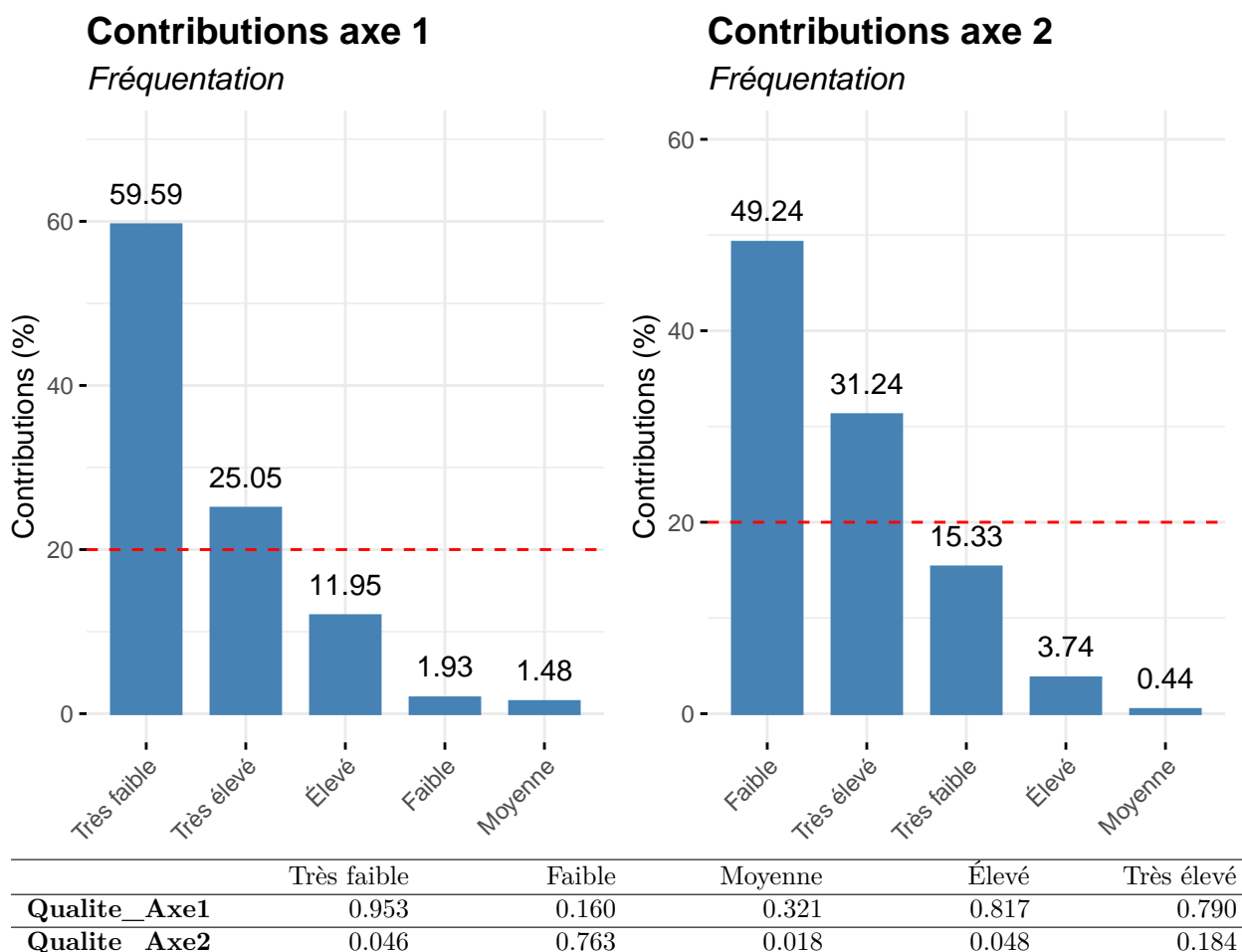
Diagramme



En examinant les données :

- Août présente une proportion élevée de jours avec une fréquentation “Très faible” (19,5%) et les proportions diminuent à mesure que l’on monte dans les niveaux de fréquentation, avec le niveau “Très élevé” le plus bas (4,5%). Cela pourrait indiquer que le mois d’août est moins populaire pour le golf, peut-être en raison de la chaleur ou des vacances d’été.
- Janvier et juillet montrent également une proportion significative de jours avec une fréquentation “Très faible”. Pour janvier, cela peut être dû au froid ou à la neige, tandis que pour juillet, d’autres facteurs saisonniers pourraient jouer.
- Mai et octobre présentent une distribution plus uniforme des niveaux de fréquentation, avec une légère tendance vers des niveaux de fréquentation “Moyenne” à “Très élevée”, ce qui peut suggérer des conditions plus favorables ou des événements qui attirent plus de joueurs.
- Décembre, février et mars montrent un équilibre relativement uniforme sur les catégories de fréquentation, bien que décembre et février aient un peu plus de jours avec une fréquentation “Faible” et “Moyenne”, ce qui peut refléter l’impact des conditions hivernales sur le jeu.

Contribution aux axes



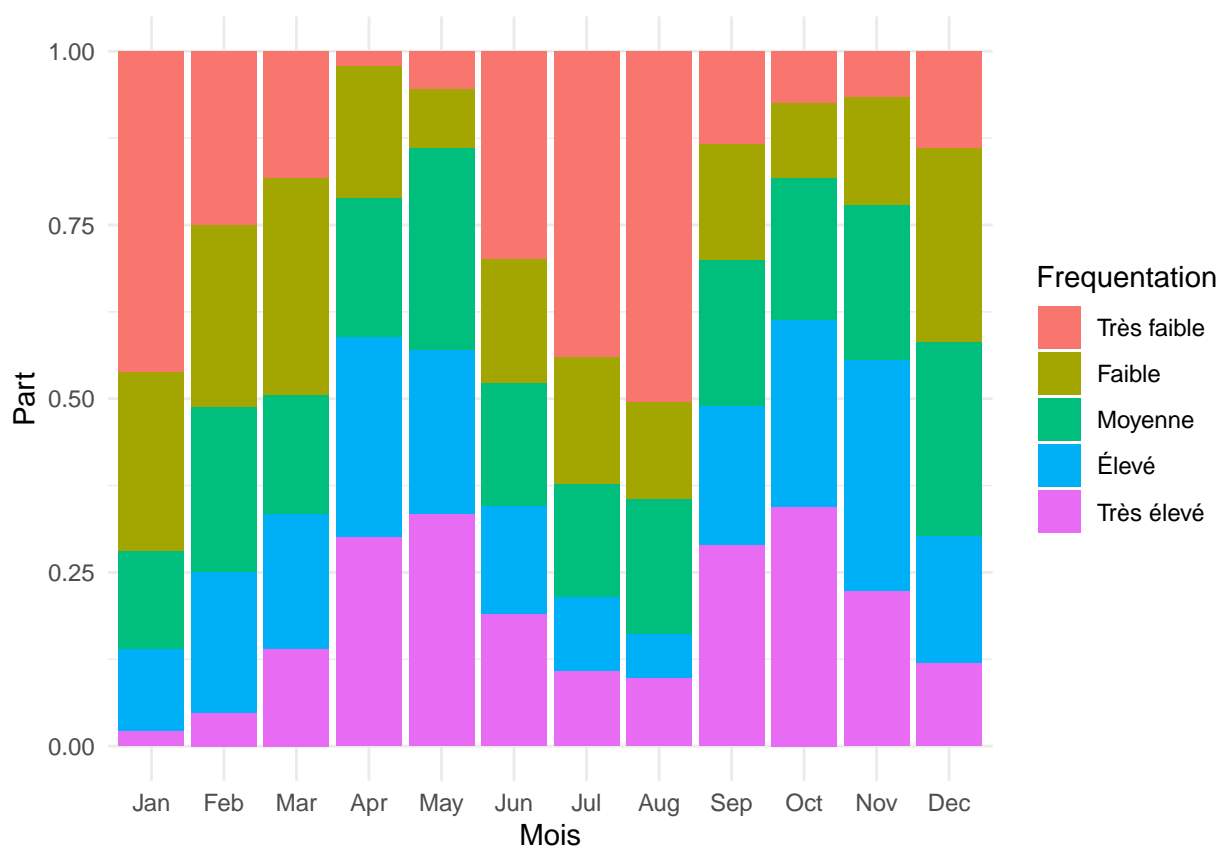
Ces contributions montrent que les niveaux extrêmes de fréquentation (“Très faible” et “Très élevé”) sont les plus distinctifs (avec des qualités de présentations très élevés) et influencent le plus les deux premiers axes de l’AFC. Cela pourrait indiquer que les facteurs qui déterminent les extrêmes de la fréquentation sont les plus importants à comprendre dans cette analyse. Les axes peuvent être interprétés comme distinguant principalement les jours de faible fréquentation des jours de forte fréquentation.

Profil ligne - Mois

Tableau de proportion

	Très faible	Faible	Moyenne	Élevé	Très élevé	Total
Apr	2.2	18.9	20.0	28.9	30.0	100
Aug	50.5	14.0	19.4	6.5	9.7	100
Dec	14.0	28.0	28.0	18.3	11.8	100
Feb	25.0	26.2	23.8	20.2	4.8	100
Jan	46.2	25.8	14.0	11.8	2.2	100
Jul	44.1	18.3	16.1	10.8	10.8	100
Jun	30.0	17.8	17.8	15.6	18.9	100
Mar	18.3	31.2	17.2	19.4	14.0	100
May	5.4	8.6	29.0	23.7	33.3	100
Nov	6.7	15.6	22.2	33.3	22.2	100
Oct	7.5	10.8	20.4	26.9	34.4	100
Sep	13.3	16.7	21.1	20.0	28.9	100
Total	22.0	19.3	20.7	19.5	18.4	100

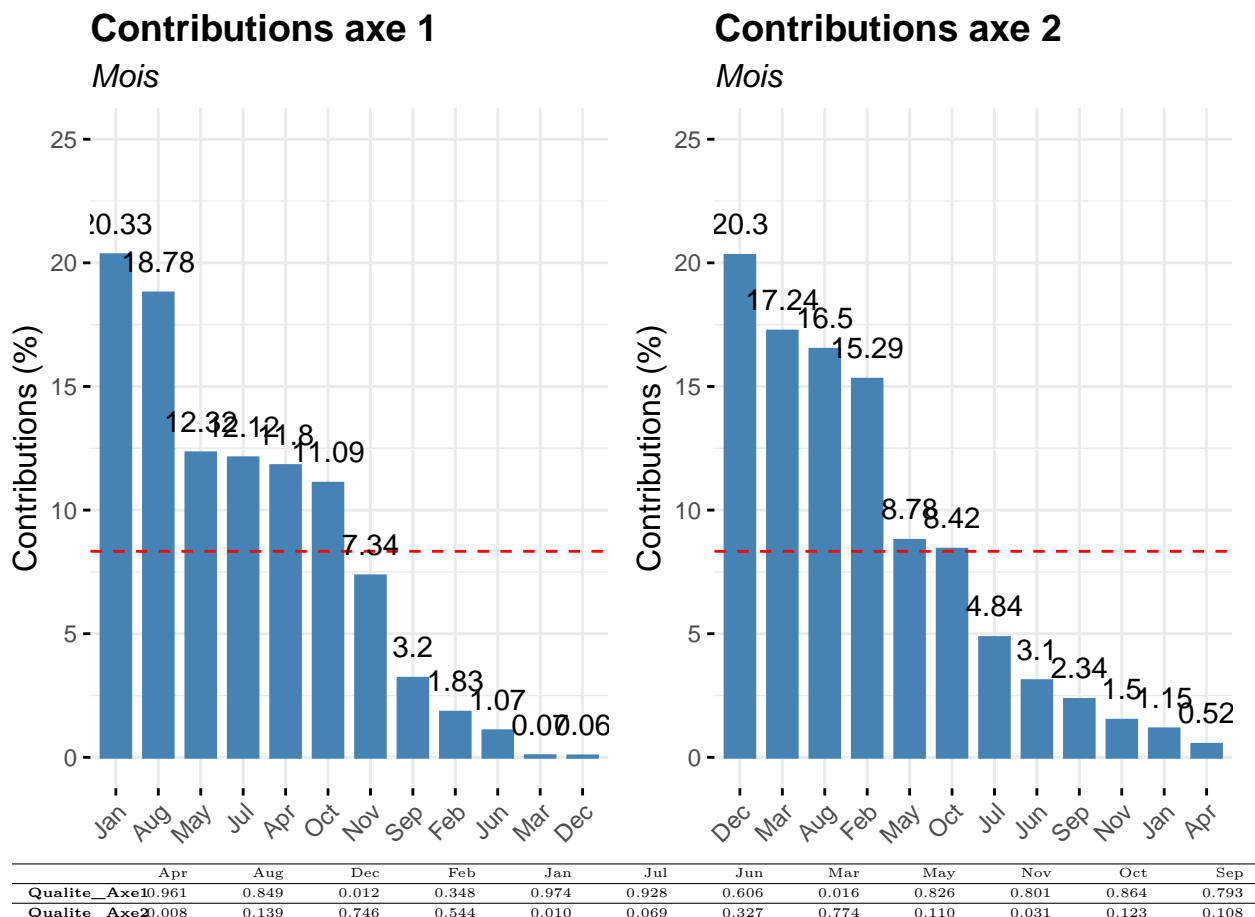
Diagramme



Les mois de janvier, juillet, août et décembre présentent une part importante de jours avec une fréquentation “Très faible” (en rouge), ce qui pourrait refléter des conditions météorologiques défavorables (températures extrêmes) ou des périodes de vacances où les joueurs sont moins disponibles.

Les mois de mai, avril, octobre et septembre sont les mois ayant la part de jours à fréquentation élevé à très élevé la plus importante. ce sont les mois où les températures sont les plus tempérées. Les joueurs de golf évitent donc les températures extrêmes.

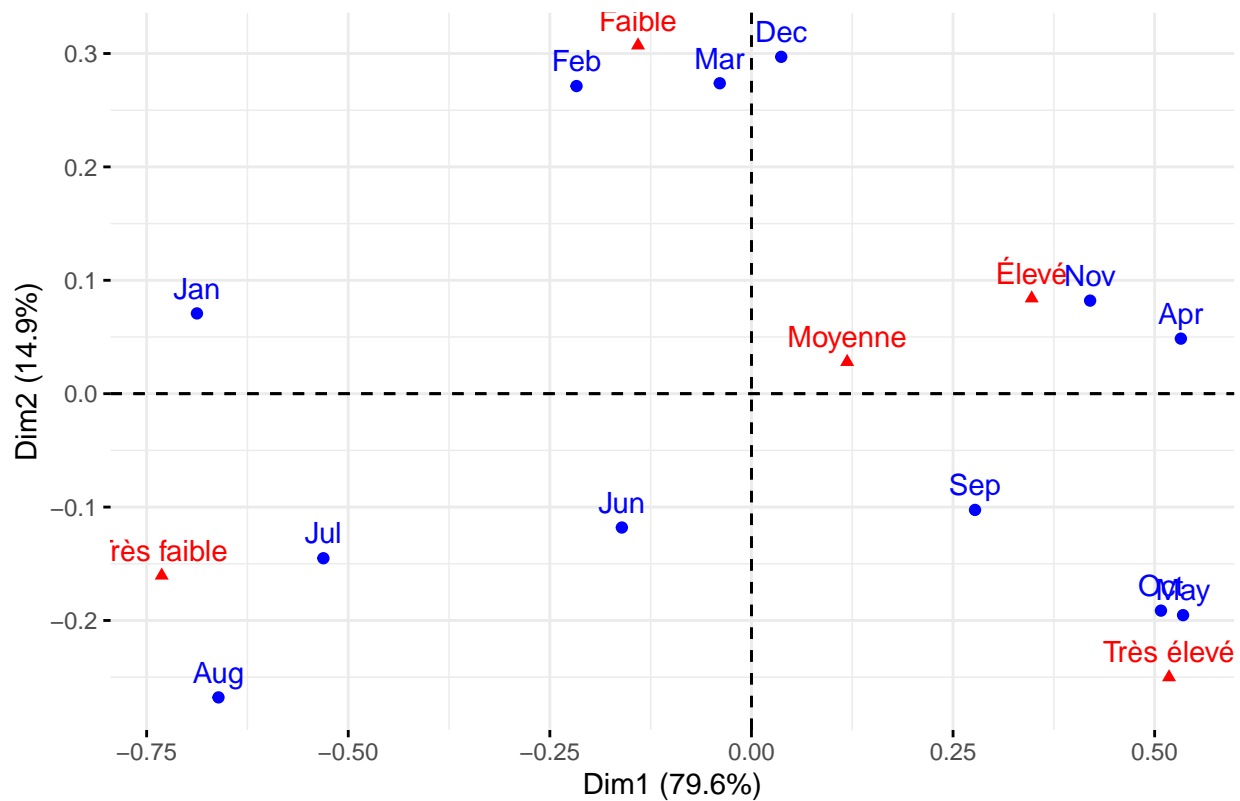
Contribution aux axes



Janvier et août sont les mois les plus contributifs à l'axe 1, avec respectivement environ 20.33% et 18.78% de contribution, ce qui peut suggérer que ces mois ont des caractéristiques distinctes qui affectent fortement la fréquentation du golf. En effet ce sont les mois où les conditions météorologiques sont les plus extrêmes. Cela indique une saisonnalité dans les habitudes de fréquentation qui peut être importante pour la planification des activités de golf.

Conclusion AFC 1

CA – Biplot



L'axe 1 oppose très distinctement les niveaux de fréquentation de très élevé à très faible avec leurs mois associés. On remarque encore la forte corrélation entre les niveaux de fréquentation extrêmes et les mois à conditions météorologiques extrêmes.

L'axe 2 semble mettre en opposition (dans une moindre mesure) des dimensions temporelles avec les mois où les températures sont les plus faibles dans la partie positives et les mois où les températures sont les plus élevées dans la partie négatives.

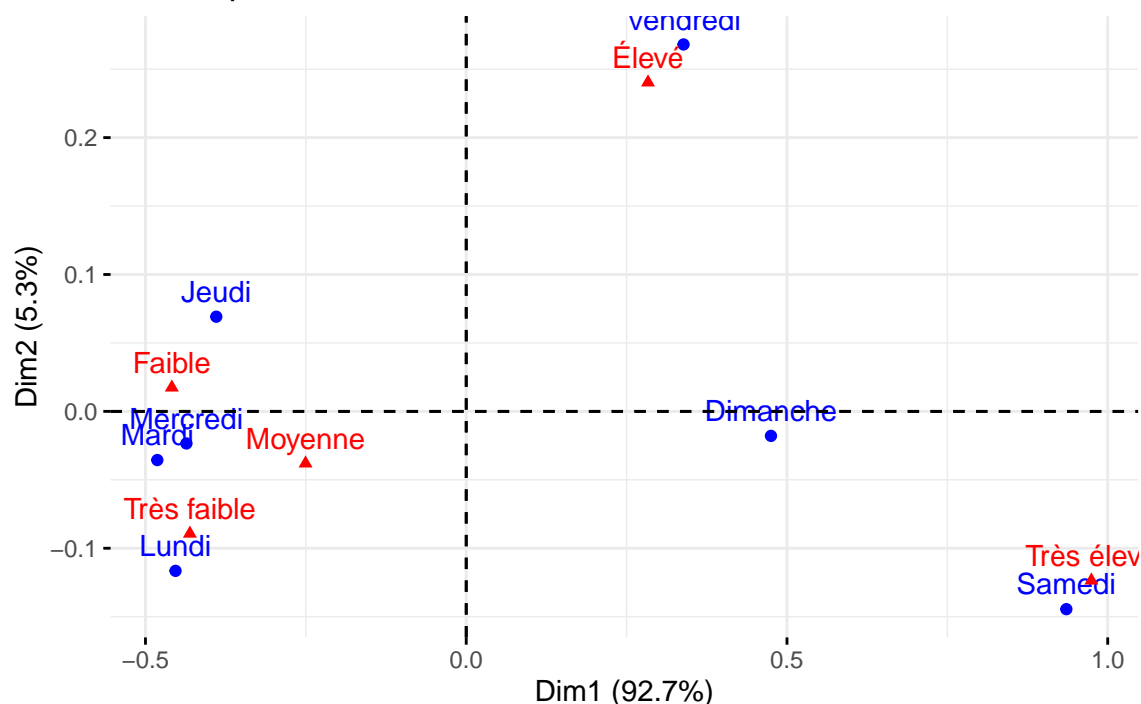
AFC - Fréquentation et Jour de la semaine

Tableau de contingence

	Très faible	Faible	Moyenne	Élevé	Très élevé	Total
Lundi	49	36	47	17	7	156
Mardi	47	45	39	20	5	156
Mercredi	52	42	32	23	7	156
Jeudi	40	44	38	27	7	156
Vendredi	19	21	26	52	39	157
Samedi	11	8	18	37	83	157
Dimanche	23	15	27	38	54	157
Total	241	211	227	214	202	1095

Conclusion AFC 2

CA – Biplot



Axe 1 (Dim1 - 92.7%) :

- Cet axe explique une très grande partie de la variance (92.7%) et semble opposer le samedi, associé à une fréquentation “Très élevée”, à tous les autres jours de la semaine. Cela suggère que le samedi est un jour particulièrement populaire pour le golf, ce qui pourrait être dû au fait que c’est le week-end et que les gens ont plus de temps libre pour des activités de loisirs.
- Le lundi est situé à l’opposé du samedi sur cet axe et est associé à une fréquentation “Très faible”, ce qui pourrait refléter une tendance pour les gens à moins fréquenter le golf en début de semaine.

Axe 2 (Dim2 - 5.3%) : - L’axe 2, bien que capturant une plus petite portion de la variance, distingue le vendredi, associé à une fréquentation “Élevée”, des autres jours de la semaine. Cela peut indiquer que le vendredi est également un jour de fréquentation relativement forte, peut-être en raison des personnes commençant leur week-end plus tôt ou participant à des événements de golf de fin de semaine. - Les autres jours de la semaine, comme le jeudi et le mercredi, semblent avoir une fréquentation “Moyenne” à “Faible”, ce qui peut indiquer une activité régulière mais pas aussi intense que le week-end.

L’analyse met en lumière une distinction claire entre les jours de la semaine en termes de fréquentation du golf. Le samedi se démarque comme étant le jour avec la plus forte fréquentation, ce qui est cohérent avec le comportement habituel des loisirs pendant le week-end.

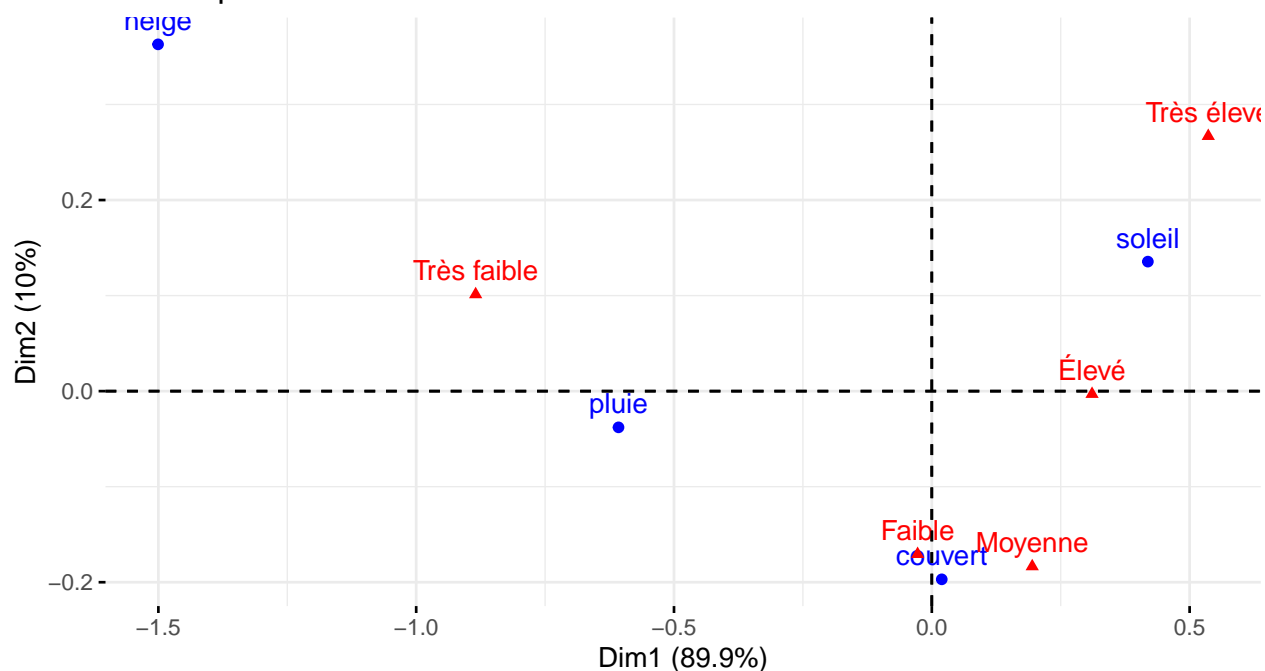
AFC - Fréquentation et Météo

Tableau de contingence

	Très faible	Faible	Moyenne	Élevé	Très élevé	Total
soleil	35	75	97	114	142	463
couvert	71	87	98	74	50	380
pluie	89	42	31	26	10	198
neige	46	7	1	0	0	54
Total	241	211	227	214	202	1095

Conclusion AFC 3

CA – Biplot



Axe 1 (Dim1 - 89.9%) : - L'axe 1 capture la grande majorité de la variance (89.9%) et met en opposition les conditions de “neige” avec les conditions de “soleil”. La neige est associée à une fréquentation “Très faible”, ce qui est cohérent avec l'idée que les conditions hivernales et enneigées dissuadent la pratique du golf. À l'opposé, le soleil est associé à des niveaux de fréquentation “Élevée” et “Très élevée”, ce qui suggère que les jours ensoleillés favorisent une plus grande affluence sur les terrains de golf. - Les conditions de “pluie” et “couvert” sont proches de l'origine sur l'axe 1, ce qui pourrait indiquer que leur effet sur la fréquentation est moins marqué par rapport à la neige ou au soleil. Néanmoins, la “pluie” est légèrement plus proche du niveau de fréquentation “Très faible”, ce qui laisse entendre que la pluie peut avoir un effet négatif sur le nombre de joueurs.

Axe 2 (Dim2 - 10%) : - L'axe 2, qui représente une plus petite portion de la variance, ne montre pas de séparation nette entre les différentes conditions météorologiques et les niveaux de fréquentation. Cependant, on peut observer que “neige” et “soleil” s'opposent à “couvert” et “pluie” ce qui peut indiquer que l'axe 2 peut séparer les conditions extrêmes et les conditions plus modérées.

Cette AFC confirme des intuitions logiques concernant l'impact de la météo sur la fréquentation des terrains de golf. Les jours de neige sont clairement associés à une fréquentation réduite, tandis que les jours ensoleillés semblent être les plus propices au golf. Les conditions de “pluie” et “couvert” semblent avoir un impact intermédiaire, suggérant que les joueurs peuvent être moins enclins à jouer par temps humide ou gris, mais pas dans la même mesure que par temps de neige.

ACM

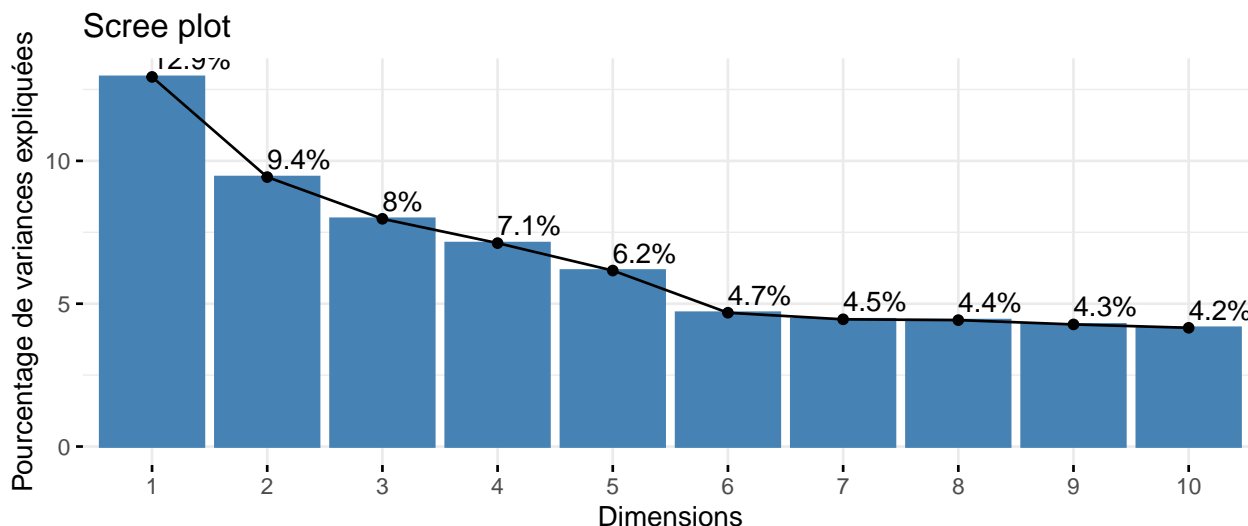
Après avoir exploré les dynamiques entre les jours de la semaine, les conditions météorologiques, les mois et les niveaux de fréquentation à travers des Analyses Factorielles des Correspondances (AFC), nous nous tournons maintenant vers une approche plus englobante : l'Analyse des Correspondances Multiples (ACM). L'ACM est particulièrement pertinente lorsque nous traitons des ensembles de données avec plusieurs variables qualitatives, car elle nous permet de visualiser et d'analyser les modèles de relations entre plus de deux catégories.

L'objectif est de tirer des conclusions plus fines sur les habitudes des joueurs et sur les facteurs qui influencent leur présence sur le terrain de golf.

Inertie des axes

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.2587700	12.9385009	12.93850
dim 2	0.1886541	9.4327072	22.37121
dim 3	0.1594219	7.9710938	30.34230
dim 4	0.1424156	7.1207810	37.46308
dim 5	0.1232080	6.1604024	43.62349
dim 6	0.0936611	4.6830570	48.30654
dim 7	0.0890631	4.4531568	52.75970
dim 8	0.0885082	4.4254090	57.18511
dim 9	0.0854968	4.2748389	61.45995
dim 10	0.0831273	4.1563662	65.61631
dim 11	0.0807698	4.0384878	69.65480
dim 12	0.0800541	4.0027072	73.65751
dim 13	0.0787821	3.9391064	77.59661
dim 14	0.0725017	3.6250828	81.22170
dim 15	0.0655353	3.2767671	84.49846
dim 16	0.0614199	3.0709938	87.56946
dim 17	0.0551588	2.7579422	90.32740
dim 18	0.0473935	2.3696729	92.69707
dim 19	0.0379187	1.8959331	94.59301
dim 20	0.0290456	1.4522801	96.04529
dim 21	0.0255111	1.2755573	97.32084
dim 22	0.0198341	0.9917066	98.31255
dim 23	0.0188853	0.9442661	99.25682
dim 24	0.0148637	0.7431834	100.00000

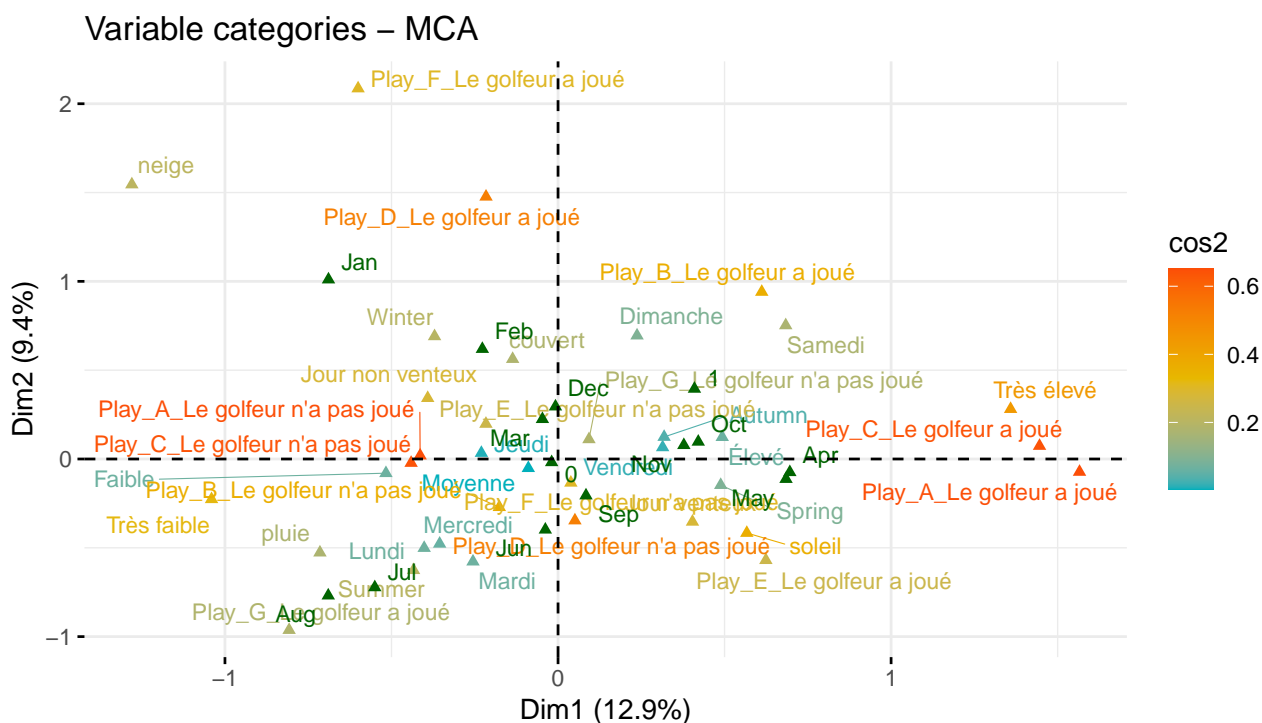
Diagramme des inerties



Ce premier graphique représente les inerties pour chaque axes factoriels.

On peut alors sélectionner ici les 3/4 premiers axes, qui explique environ 47% de l'inertie.

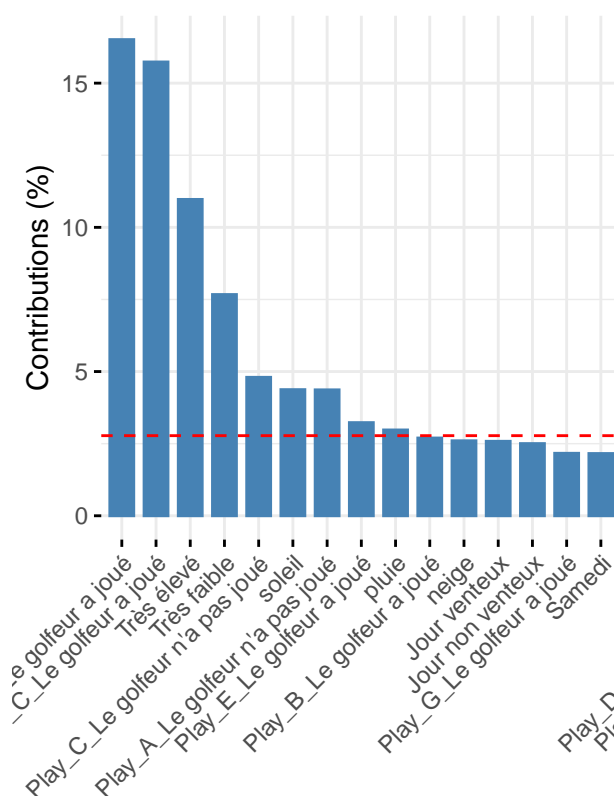
Nuages des variables - Dim 1, 2



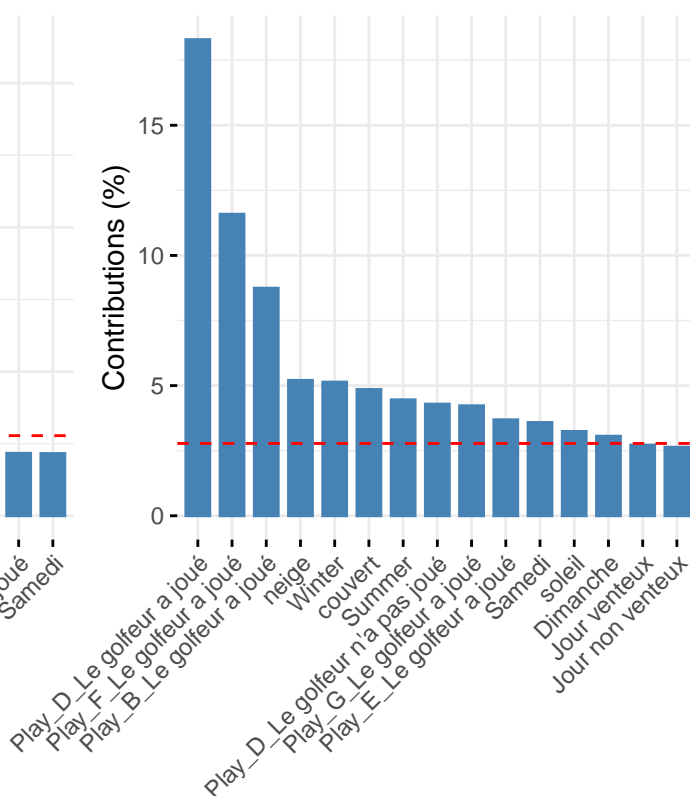
L'axe 1 mets en opposition le niveau de fréquentation de terrain de golf. La modalités élevé se retrouve positive sur cet axe alors que la modalité très faible se retrouve à l'opposé placé négativement. L'axe 1 mets également en opposition les décisions de jouer des joueurs A et C. L'axe 1 viens aussi mettre en avant l'opposition des jours de pluie avec les jours ensoleillés. L'axe 2 viens quant à elle mettre en avant les décisions de jouer des golfeurs F, D, B, et une certaine opposition des saison hiver et été. Ces variables énoncées dans cette analyse ci-dessus sont les variables les mieux représentées sur les axes 1 et 2.

Etude des contribution

Contribution of variables to Dim-1



Contribution of variables to Dim-2

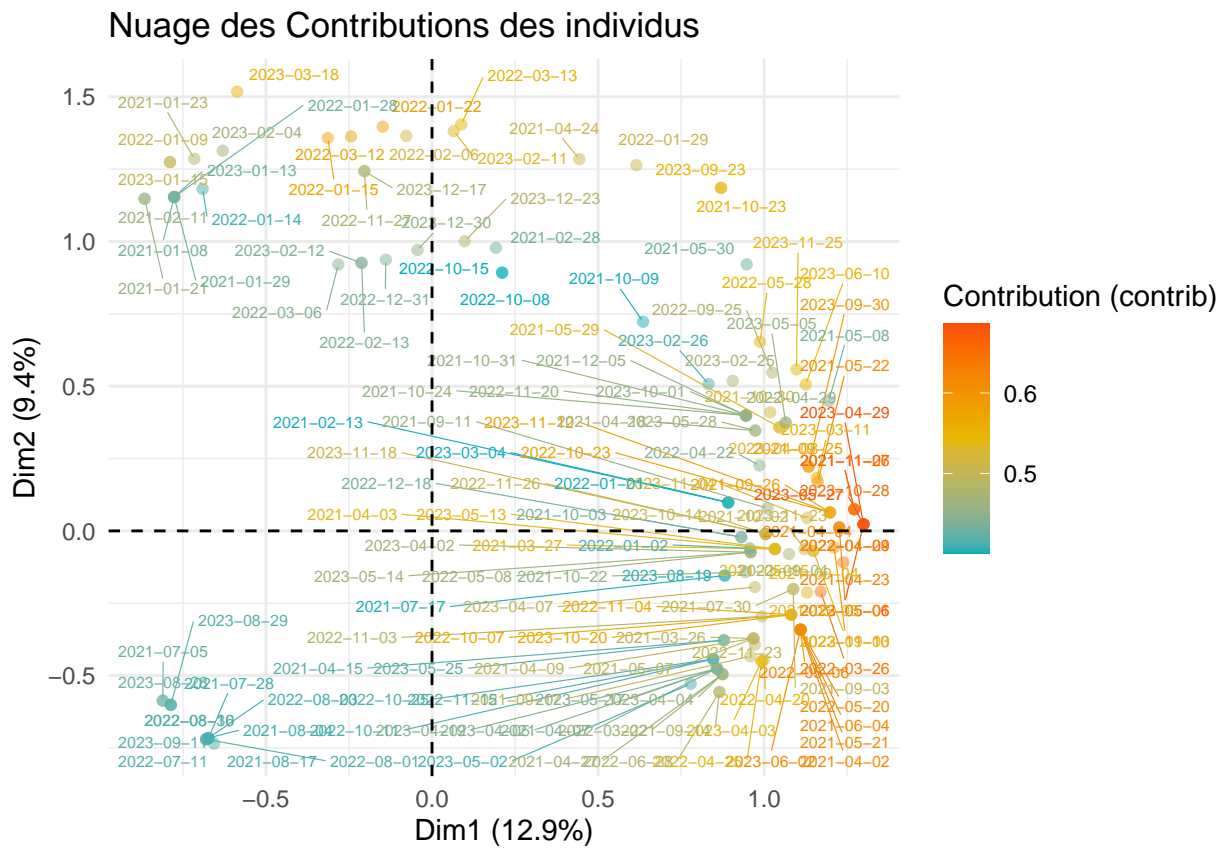


On retrouve les résultats de notre analyse plus haut. Les variables qui contribuent le plus à l'axe 1 sont les variables qui représentent la décision de jouer des golfeurs A et C, suivi de la variable "Crowdedness" avec les modalités très élevé et très faible.

Les variables qui ont le plus contribué à la construction de l'axe 2 sont les décisions de jouer des golfeurs D, F et B.

On pourra en parallèle faire une étude des individus pour confirmer notre étude. On affichera les 150 dates les plus représentatives pour avoir un graphique pertinent et suffisamment visible.

Nuage des contributions des individus



La répartition des individus semblent suivre les modalités. On peut étudier les individus “extrêmes” pour se donner un ordre d’idée de ce que représentent les axes.

	2021-01-29	2022-08-16	2023-03-18	2023-05-27
Weekday	Vendredi	Mardi	Samedi	Samedi
Month	Jan	Aug	Mar	May
Season	Winter	Summer	Winter	Spring
Holiday	0	0	0	0
Temperature	-2.0	26.5	1.4	16.9
Humidity	54	88	63	58
Windy	Jour non venteux	Jour non venteux	Jour non venteux	Jour venteux
Outlook	neige	pluie	neige	soleil
Crowdedness	0.34	0.27	0.51	0.85
Crowdedness_Cat	Très faible	Très faible	Très faible	Très élevé
Play_A	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur a joué
PlayTimeHour_A	0.0	0.0	0.0	3.6
Play_B	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur a joué	Le golfeur a joué
PlayTimeHour_B	0.0	0.0	3.1	3.2
Play_C	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur a joué
PlayTimeHour_C	0.0	0.0	0.0	4.9
Play_D	Le golfeur a joué	Le golfeur n'a pas joué	Le golfeur a joué	Le golfeur n'a pas joué
PlayTimeHour_D	3.1	0.0	3.7	0.0
Play_E	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur a joué
PlayTimeHour_E	0.0	0.0	0.0	2.8
Play_F	Le golfeur a joué	Le golfeur n'a pas joué	Le golfeur a joué	Le golfeur n'a pas joué
PlayTimeHour_F	3.3	0.0	3.9	0.0
Play_G	Le golfeur n'a pas joué	Le golfeur a joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_G	0.0	3.3	0.0	0.0

Ce tableau, prenant en compte les valeurs des dates placées à l'extrême des axes 1 et 2, vient confirmer nos interprétations précédentes : l'axe 1 traduit le caractère bondé des terrains de golf ainsi que les décisions de jouer des golfeurs A et C.

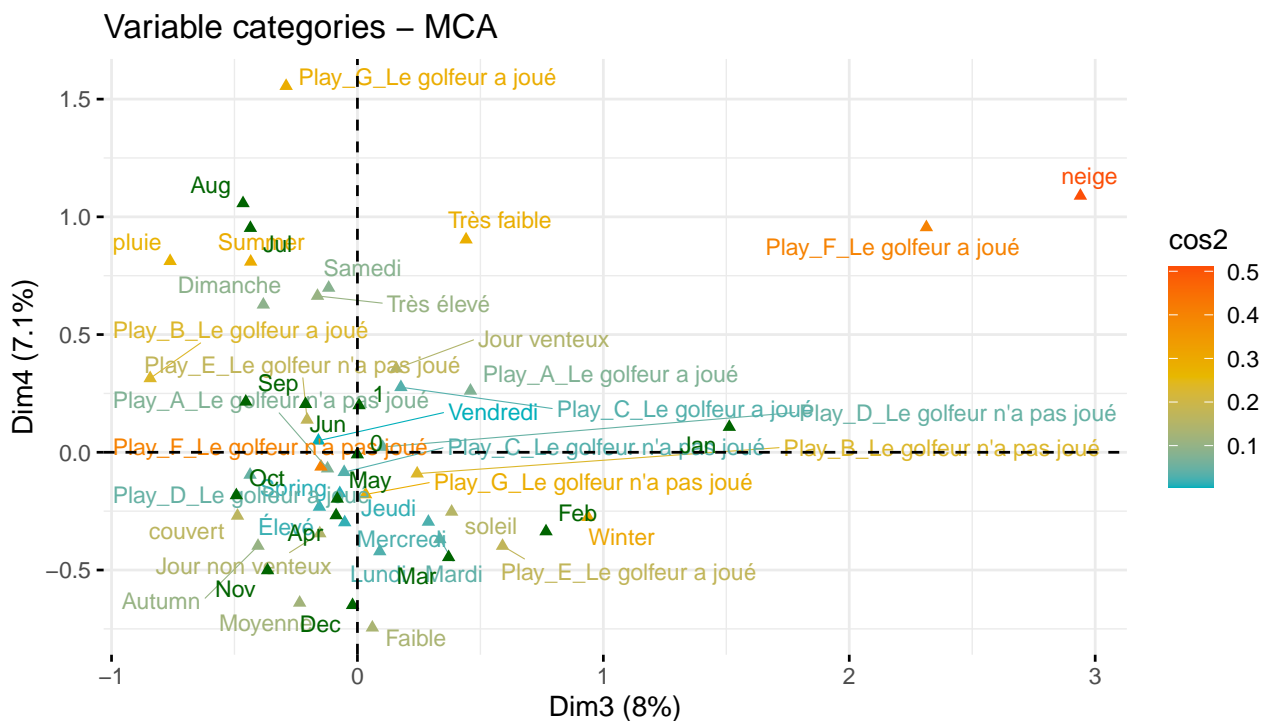
L'axe 2, mets en opposition les saisons été et hiver et les décisions de jouer des joueurs B, D et F.

Conclusion pour les deux premières axes : en se basant sur les coordonnées et les qualités de représentation des variables on peut dire que : - L'axe 1 traduit la densité de joueur présents sur les terrains de golf, si une date est placée positivement sur l'axe 1, c'est qu'à cette date, les terrains de golf étaient bondés et probablement ensoleillés.

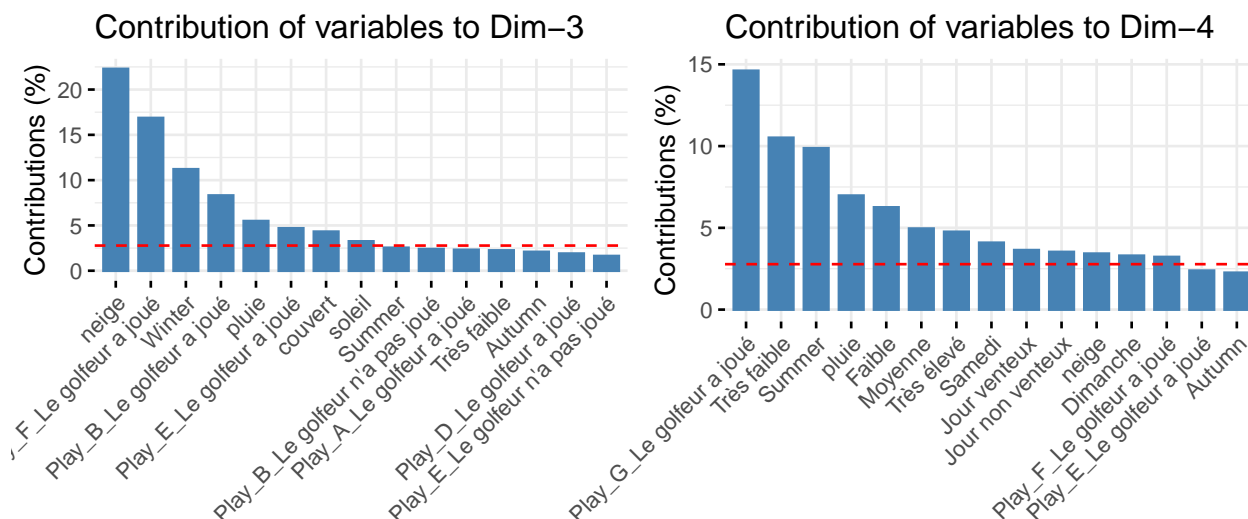
Il y a également de forte chance que ces terrains aient accueillis les joueurs A et C ce jour là. A l'inverse, une date placée négativement sur l'axe 1 reflètera un jour où les terrains de golf ne sont pas très prisés, un potentiel jour de pluie où les joueurs A et C ont préférés faire autre chose que de jouer au golf.

L'axe 2 traduit les décisions de jouer des golfeurs D, F et B, ainsi qu'une certaine opposition des saisons été/hiver. Une date placée positivement sur cet axe reflète un jour d'hiver, ayant probablement accueilli les joueurs D, F et B. A l'inverse, les dates qui y sont placées négativement seront plus des jours d'été, où ces derniers auront préféré faire autre chose que jouer une partie de golf.

Nuages des variables - Dim 3, 4



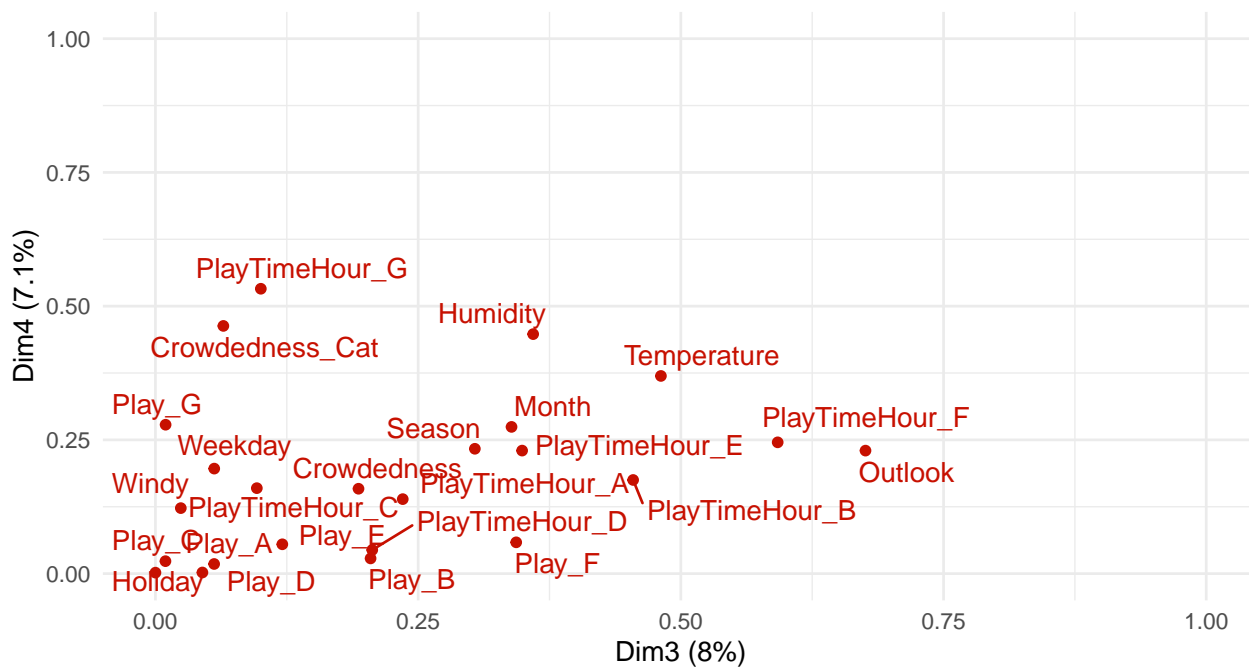
Etudes des contributions



En se basant sur la qualité de représentation des variables ainsi que leurs coordonnées on peut dire que : l'axe 3 traduit les décisions de jouer des golfeurs F et E, ainsi que les jours froid (winter) et enneigés. Ces golfeurs préfèrent donc jouer dans ces conditions météorologiques, et chaque dates placées positivement sur cette axe est synonyme d'hiver, de froid et de neige (basse température en général). L'axe 4 reflète la décision de jouer du golfeur G, une très faible densité de joueur, des jours d'été et de pluie. Comme vu précédemment dans l'ACP, le joueur G n'aime pas vraiment jouer au golf les jours où il y a beaucoup de monde. Cela l'amène à jouer les jours de pluie ou encore les jours à fortes températures ("Summer"). Les jours placés positivement sur cette axe traduisent des journées où les terrains de golfs ne sont pas très prisés, des jours d'été et de pluie. Pour fortifier l'étude, on vas pouvoir étudier les graphique des eta2 afin d'appuyer nos propos.

	dim1	dim2	dim3	dim4	dim5
PlayTimeHour_A	0.81	0.04	0.24	0.14	0.04
PlayTimeHour_C	0.81	0.04	0.10	0.16	0.27
Crowdedness	0.80	0.19	0.19	0.16	0.08
Crowdedness_Cat	0.68	0.03	0.06	0.46	0.02
Play_A	0.65	0.00	0.06	0.02	0.00
Play_C	0.64	0.00	0.01	0.02	0.08
PlayTimeHour_E	0.37	0.34	0.35	0.23	0.52
Humidity	0.34	0.38	0.36	0.45	0.11
PlayTimeHour_B	0.33	0.51	0.45	0.18	0.00
Outlook	0.32	0.35	0.68	0.23	0.33
PlayTimeHour_G	0.28	0.33	0.10	0.53	0.38
Month	0.22	0.24	0.34	0.27	0.04
Season	0.17	0.23	0.30	0.23	0.03
Windy	0.16	0.12	0.02	0.12	0.29
PlayTimeHour_F	0.16	0.53	0.59	0.25	0.01
Weekday	0.15	0.27	0.06	0.20	0.14
Play_E	0.14	0.11	0.12	0.05	0.28
Play_B	0.11	0.25	0.20	0.03	0.00
PlayTimeHour_D	0.11	0.73	0.21	0.04	0.40
Play_G	0.08	0.11	0.01	0.28	0.14
Play_F	0.02	0.28	0.34	0.06	0.00
Temperature	0.01	0.48	0.48	0.37	0.11
Play_D	0.01	0.51	0.04	0.00	0.17
Holiday	0.01	0.01	0.00	0.00	0.00

Variables – MCA

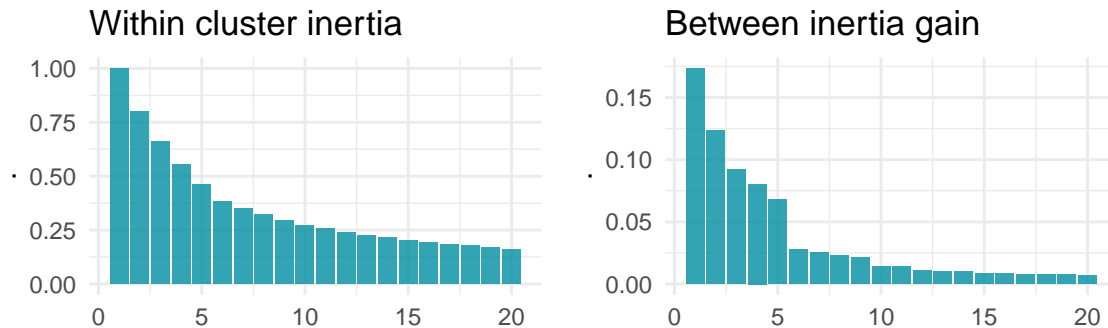


On remarque donc que la dimension 3 est très influencée par le temps de jeu de F et E, et que la dimension 4 se caractérise par le temps de jeu de G ainsi que la densité de joueur présent sur place. Ces remarques coïncident avec les études des axes 3 et 4 précédentes. La superposition de ces deux axes donne des détails clés sur les attitudes des joueurs F et G. Le fait que leurs coordonnées soient placés ainsi sur le graphique nous amène a croire qu'ils n'aiment pas jouer les jours ou il y a beaucoup de monde; car les conditions météorologiques dans lesquelles ces joueurs jouent sont assez peu communes pour jouer une partie de golf.

Clustering

Suite à notre ACM, qui a mis en lumière les interactions entre les variables qualitatives, nous allons procéder à une classification ascendante hiérarchique (CAH). C'est une méthode de clustering qui vise à construire une hiérarchie de groupes (ou clusters) parmi nos observations. La CAH va nous permettre de regrouper les individus (jours) en clusters basés sur les profils dégagés par l'ACM, identifiant ainsi des sous-groupes avec des caractéristiques similaires. Cette technique de clustering hiérarchique est particulièrement utile pour révéler la structure naturelle des données et pour identifier des groupes homogènes au sein de la population étudiée.

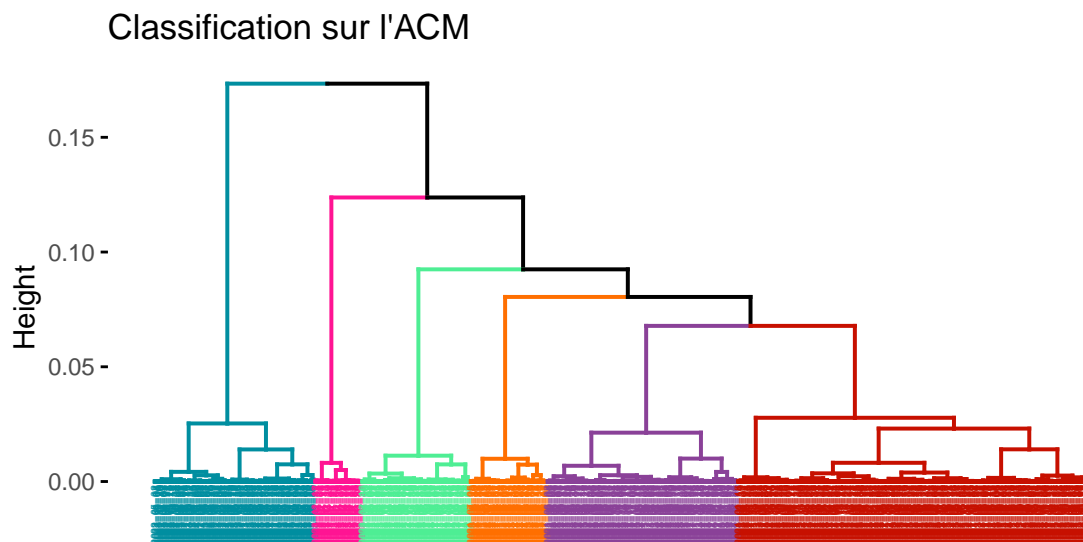
Choix des cluster



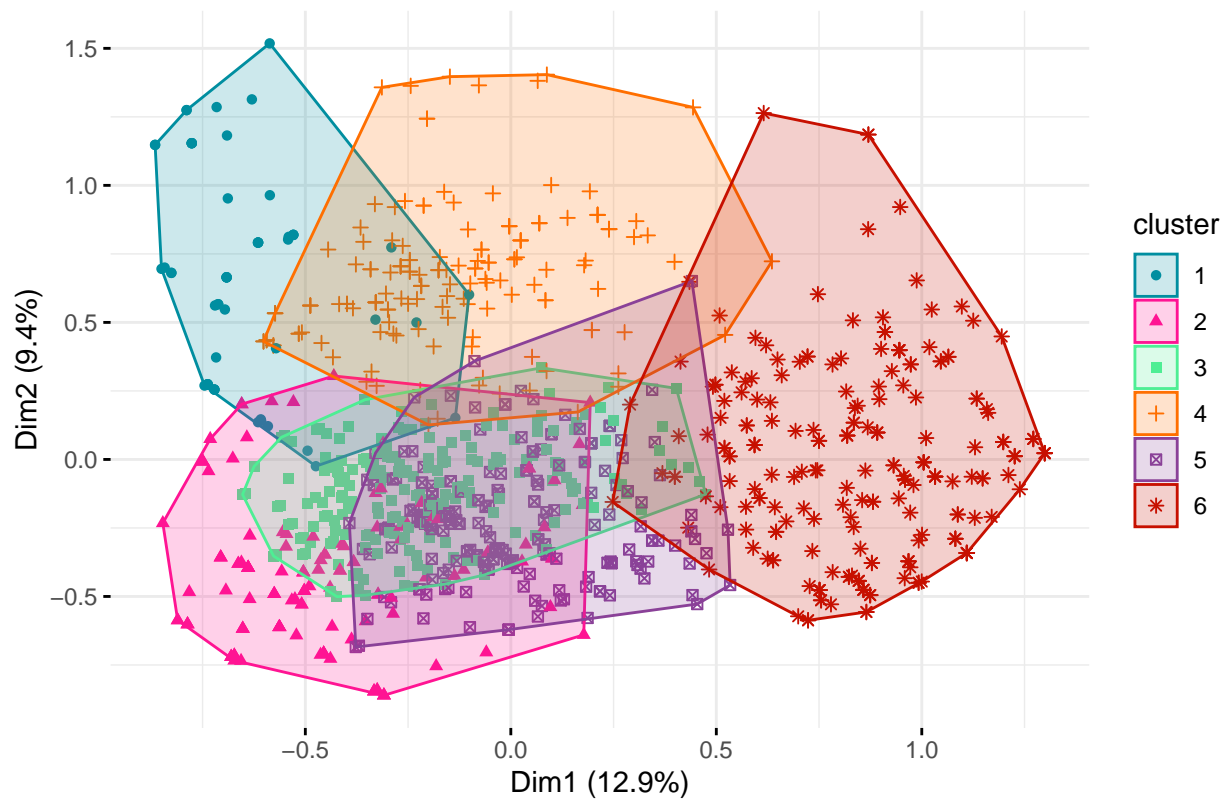
Inertie intra-cluster : Pour définir le nombre de cluster à choisir, nous recherchons un point sur le graphique où l'inertie commence à diminuer à un rythme plus lent, un point connu sous le nom de "coude". Ici les graphiques nous montre un coude entre 4 et 7 clusters.

Inertie inter-clusters : Des valeurs plus élevées indiquent un meilleur contraste entre les clusters. Comme pour l'inertie intra-cluster, nous recherchons une rupture nette où l'augmentation de l'inertie inter-clusters commence à ralentir. Ici le graphique nous montre assez distinctement une rupture du gain d'inertie à 6 clusters. Le dendrogramme ci-dessous vient confirmer nos hypothèse quant au choix du nombre de cluster. La rupture à 6 clusters est facilement visible.

Dendrogramme



Visualisation des clusters Avec consolidation



Le premier plan factoriel est le seul plan où l'on remarque assez distinctement les clusters.

Parangons

Cluster 1

	2022-01-30	2023-12-31	2021-01-10	2022-01-17	2021-02-09
Cluster 1					
Weekday	Dimanche	Dimanche	Dimanche	Lundi	Mardi
Month	Jan	Dec	Jan	Jan	Feb
Season	Winter	Winter	Winter	Winter	Winter
Holiday	0	0	0	0	0
Temperature	1.1	1.8	-0.4	-1.1	-0.4
Humidity	37	43	59	70	62
Windy	Jour venteux	Jour venteux	Jour venteux	Jour non venteux	Jour non venteux
Outlook	neige	neige	neige	neige	neige
Crowdedness	0.58	0.57	0.52	0.26	0.31
Crowdedness_Cat	Faible	Faible	Faible	Très faible	Très faible
Play_A	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_A	0	0	0	0	0
Play_B	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_B	0	0	0	0	0
Play_C	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_C	0	0	0	0	0
Play_D	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_D	0	0	0	0	0
Play_E	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_E	0	0	0	0	0
Play_F	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué
PlayTimeHour_F	3.5	3.3	3.4	3.2	3.3
Play_G	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_G	0	0	0	0	0

Le **premier cluster** regroupe des jours caractérisés par des conditions hivernales, avec des températures basses ou négatives et des conditions de neige. La présence de vent est variable, mais cela ne semble pas être un facteur distinctif pour ce cluster. La catégorie de fréquentation est généralement faible à très faible. Les jours de ce cluster tombent en hiver, majoritairement le dimanche, ce qui peut refléter une tendance à une fréquentation moindre du golf pendant les week-ends d'hiver, peut-être en raison de conditions météorologiques défavorables. De plus, il semblerait que lors de ces jours, très peu de joueurs ont joué, avec l'exception notable du golfeur F, qui a joué un nombre d'heures modeste à ces dates.

Cluster 2

	2021-06-08	2021-06-06	2023-07-07	2023-08-13	2022-08-26
Cluster 2					
Weekday	Mardi	Dimanche	Vendredi	Dimanche	Vendredi
Month	Jun	Jun	Jul	Aug	Aug
Season	Spring	Spring	Summer	Summer	Summer
Holiday	0	0	0	0	0
Temperature	21.0	21.8	22.4	26.0	26.3
Humidity	86	90	88	83	87
Windy	Jour non venteux	Jour non venteux	Jour venteux	Jour venteux	Jour venteux
Outlook	pluie	pluie	soleil	pluie	couvert
Crowdedness	0.22	0.45	0.50	0.52	0.50
Crowdedness_Cat	Très faible	Très faible	Très faible	Faible	Très faible
Play_A	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_A	0	0	0	0	0
Play_B	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_B	0	0	0	0	0
Play_C	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_C	0	0	0	0	0
Play_D	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_D	0	0	0	0	0
Play_E	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_E	0	0	0	0	0
Play_F	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_F	0	0	0	0	0
Play_G	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué
PlayTimeHour_G	3.1	3.2	3.5	3.2	3.5

Le **cluster 2** regroupe des jours qui se caractérisent par des conditions plus tempérées et humides, principalement durant les mois de printemps et d'été. Les températures sont relativement élevées, oscillant autour de 21 à 26 degrés Celsius. L'humidité est également haute, ce qui peut suggérer une atmosphère lourde ou moite. Le niveau de fréquentation est faible à très faible, indiquant que malgré des températures plus clémentes, ces jours n'attirent pas significativement plus de joueurs. Les conditions météorologiques varient de la pluie au soleil, avec des jours non venteux et venteux représentés, mais il semble que la pluie soit une condition commune au sein de ce cluster. Ceci, combiné à l'humidité élevée, pourrait dissuader les joueurs de sortir, même si les températures sont agréables. Le golfeur G semble être le seul joueur actif durant ces jours.

Cluster 3

	2021-09-16	2023-06-22	2023-06-29	2023-07-20	2022-09-30
Cluster 3					
Weekday	Jeudi	Jeudi	Jeudi	Jeudi	Vendredi
Month	Sep	Jun	Jun	Jul	Sep
Season	Summer	Summer	Summer	Summer	Autumn
Holiday	0	0	0	0	0
Temperature	18.9	15.9	21.2	22.7	21.7
Humidity	78	78	80	77	79
Windy	Jour venteux	Jour venteux	Jour venteux	Jour venteux	Jour venteux
Outlook	couvert	couvert	couvert	couvert	pluie
Crowdedness	0.55	0.53	0.60	0.52	0.55
Crowdedness_Cat	Faible	Faible	Faible	Faible	Faible
Play_A	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_A	0	0	0	0	0
Play_B	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_B	0	0	0	0	0
Play_C	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_C	0	0	0	0	0
Play_D	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_D	0	0	0	0	0
Play_E	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_E	0	0	0	0	0
Play_F	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_F	0	0	0	0	0
Play_G	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_G	0	0	0	0	0

Le **cluster 3** semble représenter des jours durant la fin du printemps et l'été, majoritairement des jeudis, avec des températures modérément chaudes allant de 15.9 à 22.7 degrés Celsius. L'humidité est relativement stable, et les conditions météorologiques sont majoritairement couvertes, avec un jour de pluie observé. La fréquentation est qualifiée de faible, ce qui peut être attribué aux jours couverts qui pourraient dissuader les joueurs de sortir, même si les températures sont agréables. Les jours sont généralement non venteux, ce qui pourrait être considéré comme des conditions favorables pour jouer au golf, cependant, cela ne se reflète pas dans la fréquentation. Aucun des joueurs (A à G) n'a joué ces jours-là, ce qui indique que malgré des conditions qui ne sont pas extrêmes, d'autres facteurs peuvent influencer la décision de ne pas jouer au golf. Le cluster 3 représente des jours où, bien que les conditions météorologiques ne soient pas défavorables, la fréquentation reste faible, peut-être en raison d'autres variables non capturées dans les données disponibles.

Cluster 4

	2022-03-20	2022-01-23	2023-01-08	2023-11-11	2021-03-13
Cluster 4					
Weekday	Dimanche	Dimanche	Dimanche	Samedi	Samedi
Month	Mar	Jan	Jan	Nov	Mar
Season	Spring	Winter	Winter	Autumn	Winter
Holiday	0	0	0	0	0
Temperature	7.2	3.0	3.3	11.4	2.7
Humidity	47	47	64	76	60
Windy	Jour non venteux	Jour non venteux	Jour non venteux	Jour non venteux	Jour non venteux
Outlook	couvert	couvert	couvert	couvert	couvert
Crowdedness	0.70	0.72	0.68	0.66	0.72
Crowdedness_Cat	Élevé	Élevé	Élevé	Moyenne	Élevé
Play_A	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_A	0	0	0	0	0
Play_B	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_B	0	0	0	0	0
Play_C	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_C	0	0	0	0	0
Play_D	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué
PlayTimeHour_D	3.2	3.3	3.2	3.5	3.7
Play_E	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_E	0	0	0	0	0
Play_F	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_F	0	0	0	0	0
Play_G	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_G	0	0	0	0	0

Le **cluster 4** se compose de jours caractérisés par une variété de conditions météorologiques couvertes, mais avec une fréquentation relativement plus élevée par rapport aux autres clusters précédemment examinés. Les températures varient de fraîches à modérées, allant de 2.7 à 11.4 degrés Celsius, et ces jours se répartissent entre les saisons de l'hiver au printemps. L'humidité est moyenne à élevée, ce qui est typique des jours couverts, et le niveau de vent est non venteux, ce qui pourrait être considéré comme confortable pour jouer au golf. Malgré le manque de soleil, la fréquentation est catégorisée de "Moyenne" à "Élevée", ce qui indique que ces conditions ne dissuadent pas nécessairement les joueurs. Le golfeur D semble jouer régulièrement durant ces jours, indépendamment des autres conditions, avec un nombre d'heures de jeu consistant.

Cluster 5

	2022-11-24	2021-05-26	2023-04-26	2021-12-01	2022-11-09
Cluster 5					
Weekday	Jeudi	Mercredi	Mercredi	Mercredi	Mercredi
Month	Nov	May	Apr	Dec	Nov
Season	Autumn	Spring	Spring	Autumn	Autumn
Holiday	0	0	0	0	0
Temperature	4.4	13.7	15.6	9.8	10.2
Humidity	59	57	55	66	51
Windy	Jour non venteux	Jour non venteux	Jour non venteux	Jour non venteux	Jour non venteux
Outlook	soleil	soleil	soleil	soleil	soleil
Crowdedness	0.51	0.69	0.68	0.70	0.74
Crowdedness_Cat	Très faible	Élevé	Élevé	Élevé	Élevé
Play_A	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_A	0	0	0	0	0
Play_B	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_B	0	0	0	0	0
Play_C	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_C	0	0	0	0	0
Play_D	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_D	0	0	0	0	0
Play_E	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué
PlayTimeHour_E	3.4	3.1	2.9	3.2	3.0
Play_F	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_F	0	0	0	0	0
Play_G	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_G	0	0	0	0	0

Le **cluster 5** regroupe des jours caractérisés par des conditions météorologiques ensoleillées, traversant différentes saisons de l'année, principalement au printemps et en automne. Les températures varient de fraîches à tempérées, avec une humidité modérée. Le vent est généralement absent, et le niveau de fréquentation est décrit comme élevé, malgré la catégorisation de “Très faible” à “Élevée”. Ces jours se produisent durant la semaine, plus précisément le mercredi, ce qui pourrait indiquer une tendance à une activité régulière de golf en milieu de semaine lorsque les conditions météorologiques sont favorables. Le golfeur E a joué durant ces jours, avec un nombre d'heures de jeu qui reste assez constant.

Cluster 6

	2023-12-02	2022-04-30	2021-10-03	2022-12-18	2021-10-02
Cluster 6					
Weekday	Samedi	Samedi	Dimanche	Dimanche	Samedi
Month	Dec	Apr	Oct	Dec	Oct
Season	Autumn	Spring	Autumn	Autumn	Autumn
Holiday	0	0	0	0	0
Temperature	4.7	12.6	14.9	10.1	14.8
Humidity	31	70	54	54	48
Windy	Jour venteux	Jour venteux	Jour venteux	Jour venteux	Jour venteux
Outlook	soleil	soleil	soleil	soleil	soleil
Crowdedness	0.73	0.84	0.87	0.85	0.82
Crowdedness_Cat	Élevé	Très élevé	Très élevé	Très élevé	Très élevé
Play_A	Le golfeur a joué	Le golfeur n'a pas joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué
PlayTimeHour_A	3.3	0.0	3.7	3.5	3.7
Play_B	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_B	0	0	0	0	0
Play_C	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué	Le golfeur a joué
PlayTimeHour_C	4.4	4.4	4.9	4.8	4.7
Play_D	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_D	0	0	0	0	0
Play_E	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_E	0	0	0	0	0
Play_F	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_F	0	0	0	0	0
Play_G	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué	Le golfeur n'a pas joué
PlayTimeHour_G	0	0	0	0	0

Le **cluster 6** semble représenter des jours de week-end avec un niveau de fréquentation élevé à très élevé et des conditions météorologiques ensoleillées. Les températures varient d'une fraîcheur automnale et printanière à un temps plus chaud, mais restent relativement modérées. Il y a une constance dans les conditions venteuses, ce qui ne semble pas dissuader la fréquentation des terrains. Ce cluster se distingue par l'activité de deux golfeurs (A et C) qui ont joué ces jours-là, indiquant une participation régulière malgré des variations de température et la présence de vent. Les autres golfeurs (B, D, E, F, G) ne semblent pas avoir été actifs durant ces jours spécifiques. En somme, le cluster 6 réunit des week-ends attrayants pour le golf avec du soleil et une bonne fréquentation, indépendamment des conditions venteuses, suggérant que ces facteurs ne sont pas un obstacle majeur à la pratique du golf pour certains joueurs réguliers.

Conclusion générales

La présente analyse a pour objectif de comprendre comment les conditions météorologiques et les caractéristiques temporelles influencent la fréquentation des terrains de golf et le temps de jeu des joueurs. À travers une série d'Analyses, nous avons pu identifier plusieurs tendances et préférences significatives.

L'ACP a révélé que les joueurs A et C sont particulièrement enclins à jouer les jours moins humides et préfèrent les jours de forte affluence, suggérant qu'ils pourraient jouer ensemble ou que leurs temps de jeu sont liés. La corrélation positive entre leurs temps de jeu renforce cette hypothèse. Le joueur B, quant à lui, semble moins sensible aux conditions météorologiques mais a tendance à jouer plus longtemps quand il y a davantage de monde sur le terrain, indiquant que la socialisation ou la compétition pourrait être un facteur motivant. Le joueur D, en revanche, est dissuadé par la chaleur et l'humidité, et préfère éviter le golf dans ces conditions, tout comme le joueur B, avec lequel il partage des temps de jeu corrélés positivement, suggérant des préférences météorologiques similaires ou des disponibilités concordantes. Le joueur E montre des préférences météorologiques comparables à celles des joueurs A et C, ce qui pourrait indiquer une tendance au jeu en groupe ou des préférences pour des conditions de jeu spécifiques. Les joueurs F et G présentent des comportements contrastés : F préfère les jours froids, tandis que G opte pour des jours chauds et humides, ce dernier évitant les jours de forte fréquentation. Cette préférence de G pour les jours moins fréquentés semble être une stratégie délibérée pour profiter d'un terrain plus calme, ce qui est corroboré par la corrélation négative entre l'humidité et la fréquentation.

Nos AFC ont approfondi la compréhension de ces comportements en mettant en lumière les liens entre les variables qualitatives telles que les conditions météorologiques, les jours de la semaine et les mois et le niveau de fréquentation. Ces analyses ont confirmé que les jours ensoleillés et en particulier durant les week-ends, attirent une fréquentation plus élevée, indépendamment des saisons. Les jours couverts, bien que moins prisés, voient tout de même une participation régulière de certains joueurs déterminés.

Les résultats de l'ACM et de la classification hiérarchique ont fourni un aperçu des groupes de jours et de joueurs avec des caractéristiques communes. Nous avons identifié des clusters distincts reflétant des jours de faible activité en raison de conditions hivernales défavorables, des jours de pluie pendant les saisons plus chaudes qui dissuadent la plupart des joueurs, et des week-ends ensoleillés où l'activité de golf est plus intense. En conclusion, notre analyse démontre un lien manifeste entre les conditions météorologiques, la fréquentation et les préférences des joueurs de golf. Les jours ensoleillés et tempérés favorisent une fréquentation plus élevée et des temps de jeu prolongés, tandis que les conditions extrêmes comme la neige ou la forte humidité ont l'effet inverse. Les jours de semaine montrent des tendances de fréquentation modérée, avec des différences marquées entre les joueurs en fonction de leurs préférences météorologiques individuelles.