



UNIVERSITÉ
CAEN
NORMANDIE

Université de Caen
Normandie



IUT Grand Ouest Normandie

Pôle de Caen

Bachelor Universitaire de Technologie
Science des Données

Deuxième année

SAÉ : Expliquer et prédire une variable quantitative à partir de plusieurs facteurs

Analyse de données du Grand Prix de golf de la Côte d'Albâtre 2024



Auteurs

Maxime GAMONDELE

Amaury PONS-FOURNIER

Rasmata SAWADOGO

Année universitaire 2024-2025

SOMMAIRE

Introduction	3
1 Analyse globale des variables explicatives	4
1.1 Analyse des index des joueurs des cinq clubs les plus représentés	4
1.2 Relation entre le score total et l'index	6
1.3 Relation entre les résultats du tour 3 et du tour 1	7
Analyse exploratoire	7
Validation du modèle	7
Conclusion sur la modélisation	9
2 Analyse des variations de l'index en fonction des résultats de chacun des tours	11
2.1 Exploration des données	11
2.2 Relations entre l'index et les scores des trois tours	11
2.3 Modélisation par régression multiple	12
Modèle de régression pour le Tour 1	14
Modèle de régression pour le Tour 2	14
Modèle de régression pour le Tour 3	14
Conclusion sur la modélisation	14
3 Analyse des variations de l'index en fonction des résultats obtenus lors des trois tours	15
3.1 Modèle de régression linéaire multiple	15
3.2 Test de la significativité du modèle par rapport au modèle constant	15
3.3 Interprétation des coefficients et tests de Student	15
3.4 Visualisation des effets avec les diagrammes de régression partiels	15
Conclusion	17
Annexe	18
Table des figures	19

Introduction

Le Grand Prix de la Côte d'Albâtre est une compétition de golf annuelle qui rassemble des joueurs amateurs expérimentés en Haute-Normandie. L'édition 2024 de cet événement se déroule sur trois jours et sur trois parcours choisis au hasard parmi les plus beaux de la région. Le responsable du Grand Prix souhaite analyser les performances des participants en exploitant les résultats et les données collectées.

À travers ce projet, nous réaliserons des études statistiques sur différentes problématiques, comme par exemple l'influence que peut avoir le club d'origine des joueurs sur leurs résultats obtenus, ou encore l'existence d'une relation entre l'index des joueurs et leurs scores obtenus durant la compétition.

L'objectif final de cette étude est de mettre en évidence des tendances et des corrélations avec les résultats du Grand Prix. Pour ce faire, nous allons utiliser différents modèles linéaires. Ce rapport détaille les étapes réalisées, les analyses effectuées ainsi que les conclusions tirées des données.

Avant de passer à la lecture du rapport, nous vous invitons à prendre connaissance de la manière dont sont attribués les points au golf. Au golf, le système de points est basé sur le nombre de coups nécessaires pour compléter chaque trou. Le but est de terminer le parcours avec le moins de coups possible. Chaque trou a un par, qui est le nombre de coups qu'un golfeur expérimenté devrait théoriquement prendre pour compléter le trou. Les scores sont généralement notés en fonction du par : un birdie signifie un coup de moins que le par, un eagle deux coups de moins, et ainsi de suite. À l'inverse, un bogey signifie un coup de plus que le par, et un double bogey deux coups de plus. Comprendre ce système de points est essentiel pour analyser les performances des joueurs et les résultats des compétitions.

De plus, au GPCA, les joueurs sont répartis dans des 'séries' en fonction de leur index. Chaque série jouera dans une ville différente durant ce week-end prolongé, comme le présente le tableau ci-dessous.

Répartition des terrains GPCA 2024							
	Grand Prix			Trophée			
	1 ^{ère} F	1 ^{ère} H	Pro	2 ^{ème} série Messieurs	3 ^{ème} série Messieurs	Seniors Mixtes	2 ^{ème} série Dames
	<= à 14,4	<= à 9,0		9,1 à 15	15,1 à 24,0	<= à 24,0	14,5 à 28,0
	Strokeplay			Strokeplay	Stableford		
Vendredi 19	Rouen Mt St Aignan	Dieppe Pourville		Le Havre/Octeville sur Mer	Etretat	Normandie Côte d'Albâtre	
Samedi 20	Normandie Côte d'Albâtre	Rouen Mt St Aignan		Dieppe Pourville	Le Havre/Octeville sur Mer	Etretat	
Dimanche 21	Etretat	Normandie Côte d'Albâtre		Rouen Mt St Aignan	Dieppe Pourville	Le Havre/Octeville sur Mer	

Remise des prix Golf d'Etretat

Notre rapport suit une progression logique pour assurer une lecture fluide et structurée. Nous débutons avec une analyse des variables explicatives, en nous intéressant aux index des joueurs des cinq clubs les plus représentés, ainsi qu'aux relations entre le score total, l'index et les résultats des tours 1 et 3. Nous poursuivons avec l'étude des variations de l'index en fonction des scores des trois tours et enfin nous terminons avec l'étude des variations de l'index en fonction des scores des trois tours .

1 - Analyse globale des variables explicatives

1.1 Analyse des index des joueurs des cinq clubs les plus représentés

Sur ce premier graphique (cf. Figure 1), nous cherchons à savoir si certains clubs ont présenté des joueurs ayant obtenu, au final, de meilleurs résultats. Dans notre contexte, nous avons décidé d'utiliser des boxplots parallèles mettant en avant la distribution des scores totaux de chaque joueur par club. Nous nous sommes concentrés uniquement sur les cinq clubs les plus représentés, à savoir Le Havre, Dieppe, Rouen, Vaudreuil et Étretat.

À première vue, il semble que le club d'Étretat ait proposé de meilleurs joueurs. Pour vérifier cette observation, nous décidons de mettre en place un test de Fisher global. Toutefois, nous devons d'abord vérifier les hypothèses d'échantillonnage.

- **H3.1 : Indépendance des observations**

- Nous supposons ici que les variables sont indépendantes entre elles.

- **H3.2 : Homoscédasticité des distributions**

- Pour vérifier cette hypothèse, nous avons réalisé un test de Levene.

- La p-value obtenue est supérieure à 0,05, ce qui nous conduit à conserver l'hypothèse nulle. Cela confirme que nos distributions respectent bien le critère d'homoscédasticité (cf. tableau 1).

Df	F value	Pr(>F)
4	0.56	0.69
46		

TABLE 1 – Test de Levene pour l'homogénéité des variances

- **H3.3 : Normalité des distributions**

- Nous avons utilisé un test de Shapiro-Wilk pour évaluer le caractère gaussien des distributions.

- Ici, toutes les p-values sont supérieures à 0,05, ce qui nous permet de confirmer l'hypothèse de normalité (cf. tableau 2).

Club	p_value_shapiro
DIEPPE	0.93
ETRETAT	0.49
LE HAVRE	0.89
ROUEN	0.26
VAUDREUIL	0.51

TABLE 2 – Test de Shapiro-Wilk par club

Après avoir vérifié nos hypothèses d'échantillonnage, nous pouvons maintenant passer au test de Fisher. Comme le montre notre sortie graphique, la p-value obtenue est de 0,64 et la statistique de test F est proche de 1. Cela signifie que nous conservons l'hypothèse nulle, selon laquelle le modèle vide est valide. Autrement dit, il n'existe pas de différence significative entre les moyennes des scores totaux des différents clubs.

Dans notre contexte, nous pouvons en conclure qu'Étretat, ou tout autre club, n'a pas présenté de joueurs ayant obtenu des résultats significativement meilleurs. Le fait que les clubs d'Étretat et de Vaudreuil occupent des positions extrêmes est probablement dû à l'aléa.

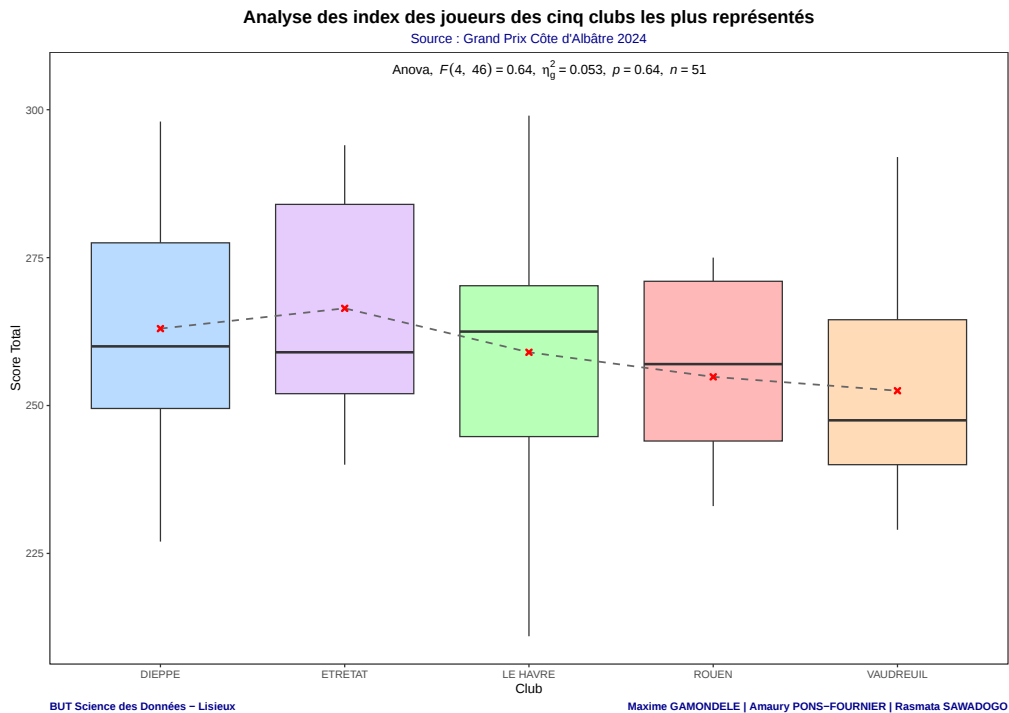


Fig. 1 - Analyse des index des joueurs des cinq clubs les plus représentés

1.2 Relation entre le score total et l'index

Le graphique ci-dessous (cf. Figure 2) illustre l'association entre l'index d'un joueur et le résultat final du joueur. On observe une association linéaire, avec une droite de régression ascendante, ce qui témoigne de la relation positive entre l'index et le score : plus l'index d'un joueur est élevé, plus son score final tend à être élevé. De plus, la p-value très inférieure à 0,05 indique que la relation observée est statistiquement significative, ce qui confirme que l'index est un facteur important pour prédire la performance des joueurs dans cette compétition.

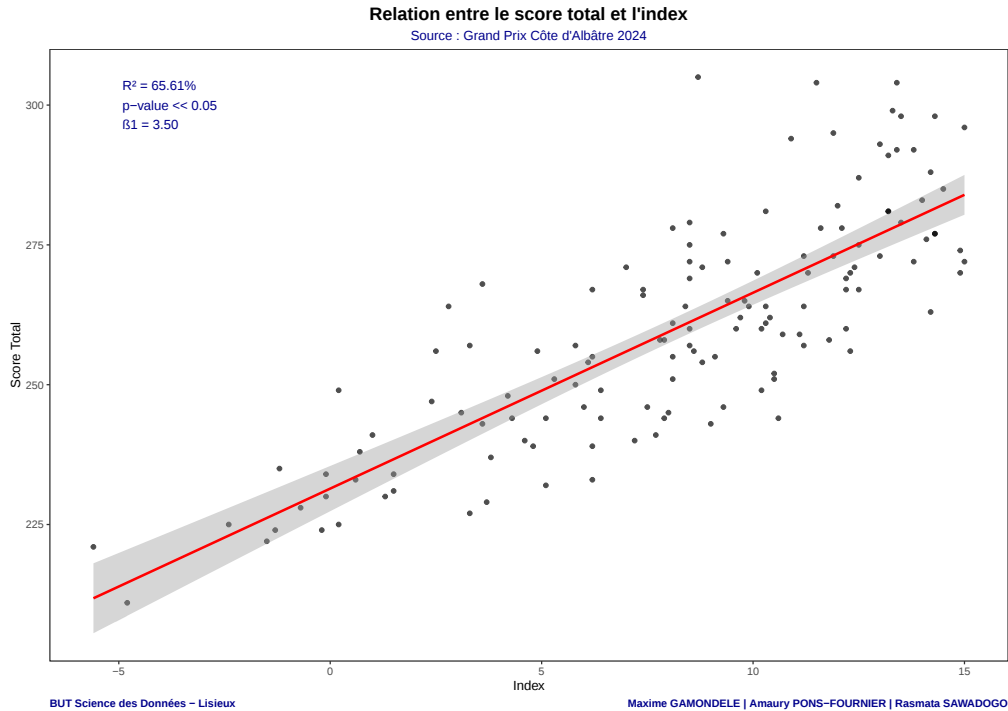


Fig. 2 - Relation entre le score total et l'index

La droite des moindres carrés est définie par l'équation suivante :

$$\hat{m}(\text{Index}) = 231.42 + 3.50 \times \text{Index} \quad (1)$$

Ce modèle révèle un coefficient de détermination R^2 de 0,6561. Cela signifie que près de 65,61% de la variation du score total peut être expliquée par l'index du joueur, indiquant ainsi une relation significative entre ces deux variables. Le coefficient associé à l'index est de 3,50, ce qui implique que pour chaque augmentation d'une unité de l'index, on s'attend à une augmentation moyenne de 3,50 points du score total.

1.3 Relation entre les résultats du tour 3 et du tour 1

Analyse exploratoire

À travers cette partie, nous cherchons à déterminer s'il existe une association entre le résultat obtenu au Tour 1 et celui obtenu au Tour 3. Pour ce faire, nous proposons de réaliser un nuage de dispersion et d'essayer de modéliser une éventuelle association.

En étudiant notre nuage de dispersion, nous avons pu remarquer la présence de valeurs atypiques qui pourraient amener une forte modification de la structure du modèle (cf. Figure 3).

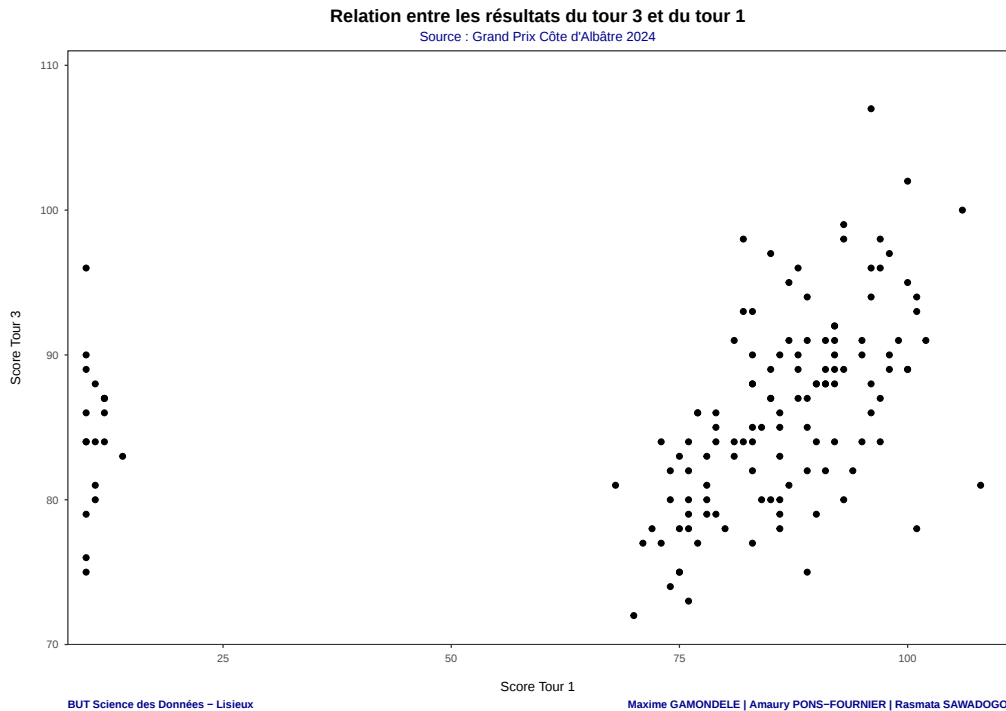


Fig. 3 - Relation entre les résultats du tour 3 et du tour 1

Validation du modèle

En raison de ces valeurs atypiques, nous décidons de tester notre modèle de régression linéaire simple avec l'ensemble de nos observations, en nous assurant que les hypothèses inhérentes soient vérifiées. Pour ce faire, nous nous concentrons sur le graphique ci-dessous (cf. Figure 4).

Premièrement, nous étudions l'hypothèse de linéarité (graphique : 1x1). Si l'hypothèse était valide, nous nous attendrions à observer les points s'aligner autour d'une droite sur $y = 0$. Or, ce n'est pas le cas ici, on semble clairement observer un "coude" entre deux nuages de dispersion.

Deuxièmement, nous cherchons à savoir si notre distribution est normalement distribuée. Ici, sur notre QQ-plot (graphique : 1x2), on semble observer que les points s'alignent sur la première bissectrice. Cela signifie que nous sommes probablement dans un cadre gaussien.

Troisièmement, nous cherchons à vérifier l'hypothèse d'homoscédasticité. Sur notre nuage de dispersion (graphique : 1x3), on observe, encore une fois, un coude et une courbe dans le plus gros nuage. Nous rejetons donc l'hypothèse d'homoscédasticité.

Quatrièmement, nous cherchons à détecter la présence d'une ou plusieurs observations influentes. Pour cela, nous regardons notre graphique des distances de Cook (graphique : 2x1). On semble observer trois points influents qui modifient la structure de notre modèle.

Dernièrement, nous allons nous concentrer sur la détection de X et/ou Y-outliers (graphique : 2x2). Sur ce dernier, on distingue clairement la présence de X-outliers sur la droite et un potentiel Y-outlier en prenant la bande $(-3, 3)$.

En conclusion, nous ne pouvons réaliser un modèle linéaire sans corriger notre échantillon c'est pour cela que nous proposons une deuxième modélisation que nous allons d'abord vérifier.

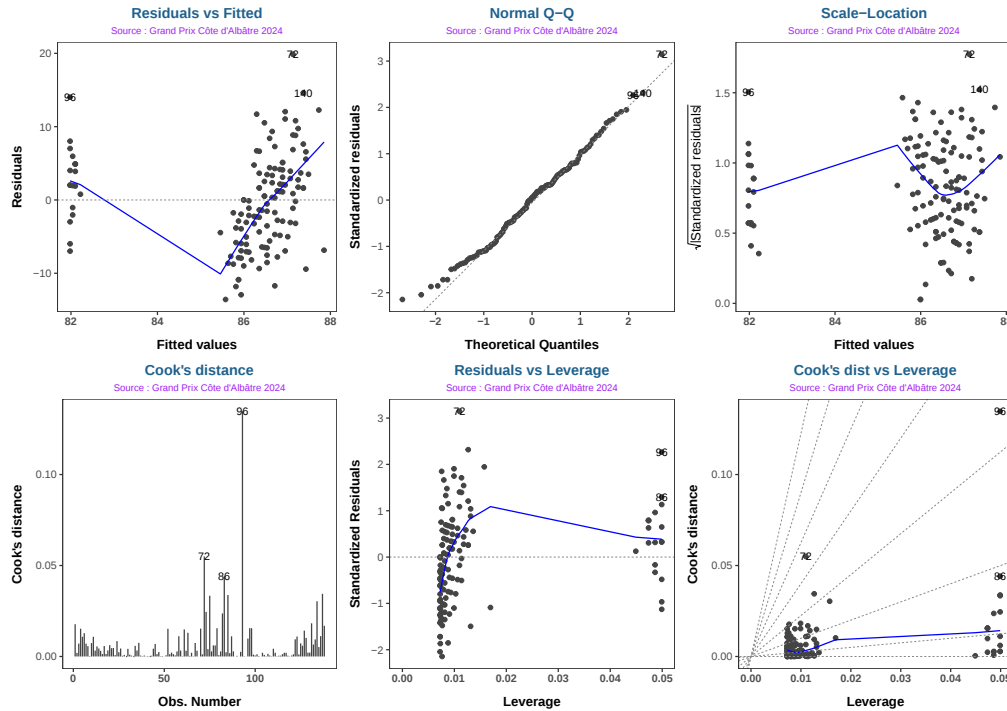


Fig. 4 - Relation entre les résultats du tour 3 et du tour 1

Comme évoqué précédemment, nous avons décidé de modifier notre échantillon en supprimant nos valeurs atypiques, notamment les X-Outliers. En refaisant nos tests pour valider notre modèle, nous pouvons observer que cette fois-ci, nos critères semblent être réunis pour valider ce dernier (cf. Figure 5).

Néanmoins, nous observons tout de même que l'individu 109 semble être un outlier qui apparaît dans plusieurs graphiques, notamment dans celui des distances de Cook (graphique : 2x1) et dans le graphique des outliers (graphique : 2x2).

En analysant les performances du joueur M. Mouty du club de Vaudreuil (individu 109), nous constatons effectivement que son score au Tour 1 semble supérieur à la moyenne des joueurs ayant un index similaire. Toutefois, son score au Tour 3 semble être dans la moyenne des joueurs de son niveau. Nous supposons donc qu'il a sous-performé lors du premier tour et décidons de le garder dans notre modélisation.

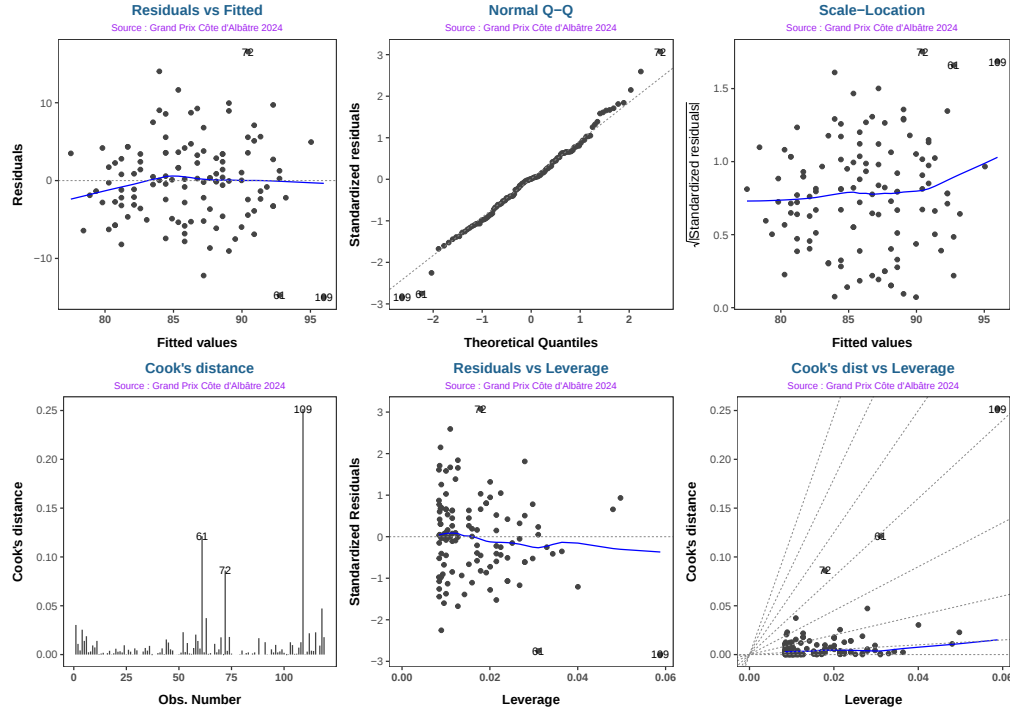


Fig. 5 - Relation entre les résultats du tour 3 et du tour 1

Conclusion sur la modélisation

Maintenant que nous avons pu valider notre modèle, nous passons à son étude. Sur notre graphique (cf. Figure 6), nous pouvons observer une association linéaire positive entre les scores du Tour 1 et ceux du Tour 3. Cela signifie que plus le score au Tour 1 est faible, plus le score au Tour 3 le sera également.

Notre modèle explique 35,45 % de la variabilité des scores obtenus au Tour 3. De plus, avec une p-value proche de 0 et une statistique de Fisher de 259, nous concluons que notre modèle est significativement meilleur que le modèle constant. Cela signifie que les coefficients de $\tilde{\beta}$ (ici $\hat{\beta}_1$ uniquement) sont significativement différents de 0.

La droite des moindres carrés est définie par l'équation suivante :

$$\hat{m}(\text{Tour1}) = 46.11 + 0.46 \times \text{Tour1} \quad (2)$$

Autrement dit, pour une diminution de 1 point au Tour 1, on observera une diminution de 0,46 points au Tour 3. Nous parlons ici de diminution, car au golf, l'objectif est d'obtenir le score le plus bas possible.

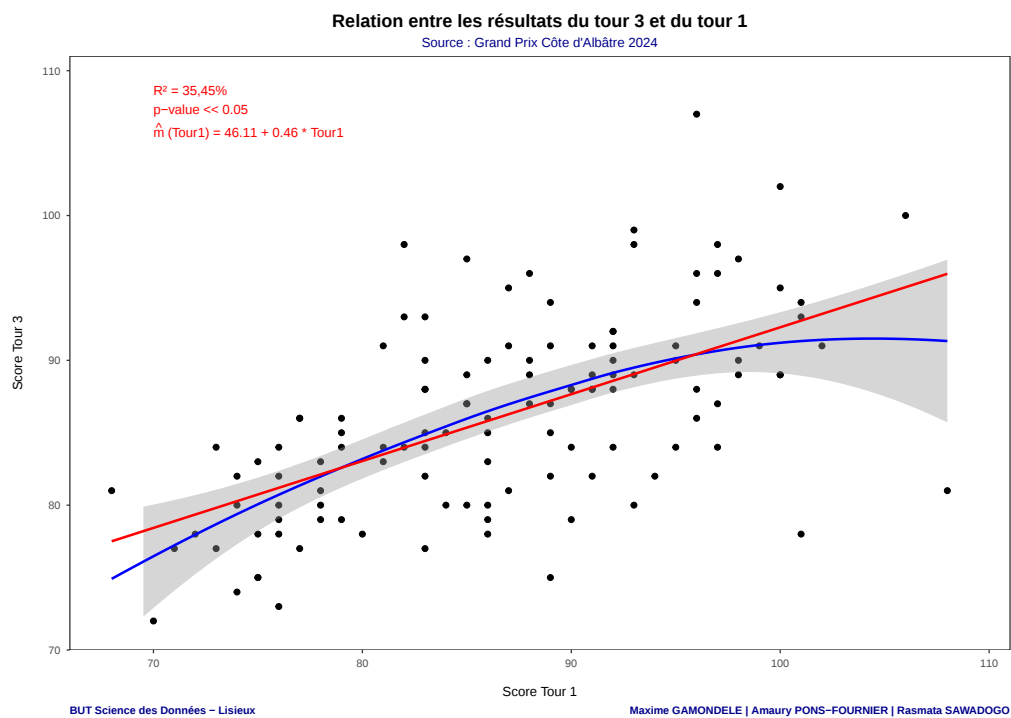


Fig. 6 - Relation entre les résultats du tour 3 et du tour 1

2 - Analyse des variations de l'index en fonction des résultats de chacun des tours

2.1 Exploration des données

Avant de passer à la modélisation, nous avons exploré les données disponibles concernant les scores des joueurs pour chaque tour et leur index. Cette phase d'exploration nous a permis d'identifier des tendances générales et de vérifier la qualité des données.

Le graphique ci-dessous (cf. Figure 7) montre les relations entre les résultats des trois tours (Tour1, Tour2, Tour3) et l'index des joueurs, ainsi que les différences par Série. On observe que le Tour2 présente la corrélation la plus forte avec l'index (0,807), ce qui en fait le tour le plus prédictif de l'index des joueurs, tandis que le Tour1 (0,769) et le Tour3 (0,654) sont également liés à l'index, mais moins fortement que le Tour2. On remarque également une différence marquée par Série : les joueurs de la Série 1 ont des indices généralement plus bas, tandis que ceux de la Série 2 ont des indices plus élevés. Cette séparation est visible dans tous les tours et souligne l'importance de prendre en compte la Série dans l'analyse.

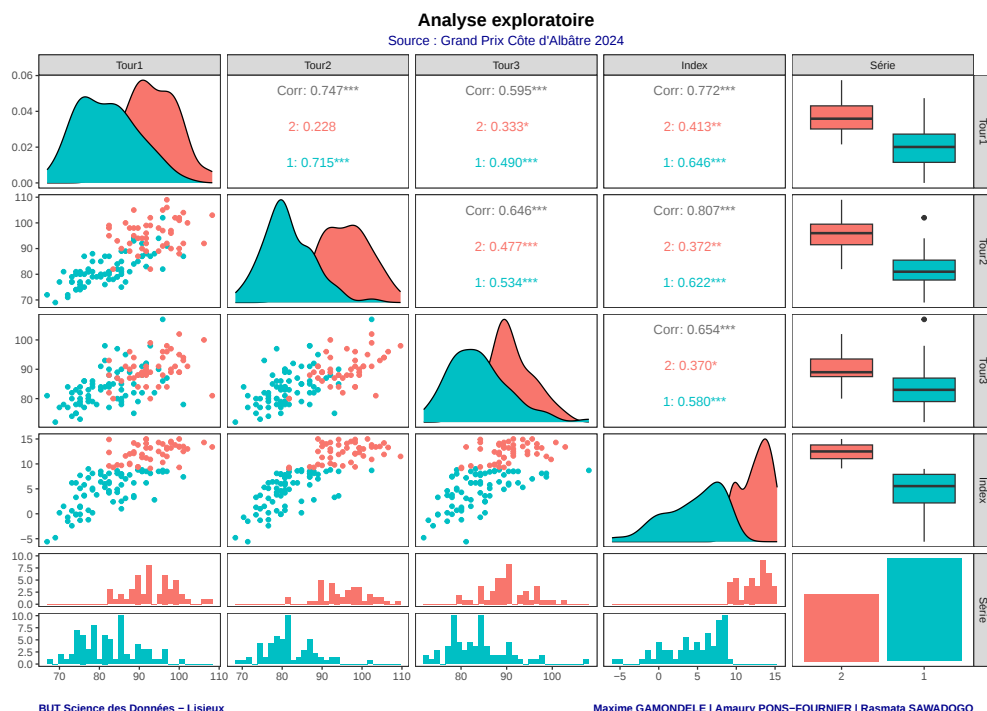


Fig. 7 - Extrait du jeu de données

2.2 Relations entre l'index et les scores des trois tours

Pour mieux comprendre l'impact des performances des joueurs sur leur index, nous avons utilisé des visualisations graphiques pour observer la relation entre l'index et les scores des trois tours (Tour 1, Tour 2, Tour 3). Le graphique ci-dessous (cf. Figure 8) illustre clairement ces relations avec une distinction par Série.

On observe que les régressions sont linéaires et positives, ce qui signifie qu'à mesure que les scores augmentent, l'index des joueurs tend également à augmenter de manière régulière. Les

pentés des droites de régression sont plus marquées pour la **Série 1**, ce qui suggère que l'impact des scores sur l'index est plus prononcé pour les joueurs de cette série. De plus, les droites de régression pour chaque série ne sont pas parallèles, ce qui indique une interaction entre les scores et la série. Cette observation nous amène à envisager un modèle qui intègre cette interaction, afin d'expliquer plus précisément les variations de l'index.

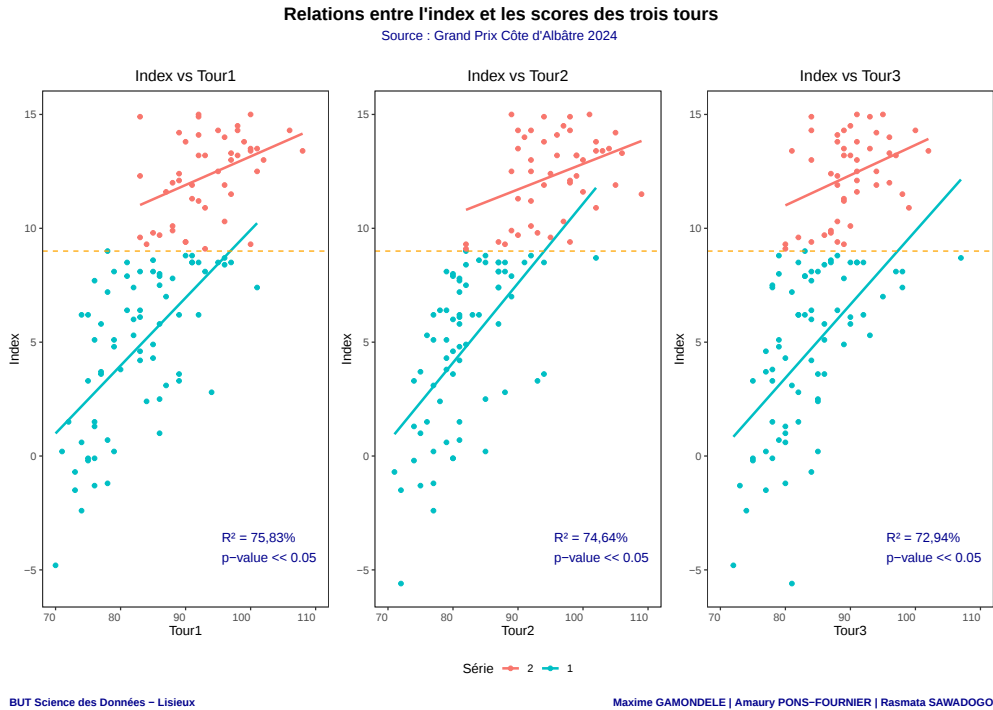


Fig. 8 - Relations entre l'index et les scores des trois tours

2.3 Modélisation par régression multiple

Afin d'analyser l'effet des performances des joueurs sur leur index, nous avons adopté une approche de régression multiple avec interaction. Dans notre modèle, la Série 2 est choisie comme modalité de référence afin de mettre en évidence que les joueurs de la Série 1 ont un index inférieur à ceux de la Série 2, ce qui souligne le fait qu'ils soient meilleurs. Pour tester l'intérêt de l'interaction entre $Tour(i)$ et $Série1$, nous réalisons un test de Student sur le paramètre correspondant. Les hypothèses en présence sont :

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_3 \neq 0 \quad (3)$$

Les p-values (cf. sorties textuelles des modèles) obtenues pour l'interaction $Tour(i) : Série1$ dans les trois modèles sont :

- **Tour1 : Série1** $\Rightarrow$ p-value = **0,0069**
- **Tour2 : Série1** $\Rightarrow$ p-value = **0,00096**
- **Tour3 : Série1** $\Rightarrow$ p-value = **0,0313**

Avec un seuil de signification de 5%, ces valeurs nous permettent de rejeter l'hypothèse H_0 pour chaque interaction. Nous concluons donc que l'interaction entre Tour(i) et Série1 a un effet significatif sur l'index.

L'équation mathématique du modèle de régression s'écrit comme suit :

$$m(\text{Tour}(i), \text{Serie}) = \beta_0 + \beta_1 \times \text{Tour}(i) + \beta_2 \times \mathbb{1}_{\text{Serie1}}(\text{Serie}) + \beta_3 \times \mathbb{1}_{\text{Serie1}}(\text{Serie}) \times \text{Tour}(i) + \epsilon \quad (4)$$

où :

- **Index** représente l'index des joueurs, qui est la variable à expliquer.
- **Tour(i)** désigne le score obtenu par le joueur au tour i .
- $\mathbb{1}_{\text{Serie1}}(\text{Serie})$ est une variable indicatrice prenant la valeur 1 si le joueur appartient à la Série 1 et 0 sinon.
- $\beta_0, \beta_1, \beta_2, \beta_3$ sont les paramètres du modèle à estimer.
- ϵ représente l'erreur résiduelle du modèle.

```
Call:
lm(formula = Index ~ Tour1 * Série, data = data4)

Residuals:
    Min       1Q   Median       3Q      Max
-5.5827 -1.4772  0.2195  1.5583  5.8101

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.61688    5.53375   0.111 0.911433
Tour1         0.12538    0.05908   2.122 0.035975 *
Série1       -22.44307    6.41880  -3.496 0.000671 ***
Tour1:Série1   0.19534    0.07096   2.753 0.006869 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.419 on 115 degrees of freedom
Multiple R-squared:  0.7583,    Adjusted R-squared:  0.752
F-statistic: 120.3 on 3 and 115 DF,  p-value: < 2.2e-16
```

Fig. 9 - Sortie textuelle modèle Tour1

```
Call:
lm(formula = Index ~ Tour2 * Série, data = data4)

Residuals:
    Min       1Q   Median       3Q      Max
-6.5956 -1.7021  0.3484  1.7663  4.4870

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.61837    5.74546   0.282 0.778696
Tour2         0.11209    0.05997   1.869 0.064147 .
Série1       -27.54420    6.99635  -3.937 0.000142 ***
Tour2:Série1   0.26182    0.07728   3.388 0.000965 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.478 on 115 degrees of freedom
Multiple R-squared:  0.7464,    Adjusted R-squared:  0.7398
F-statistic: 112.8 on 3 and 115 DF,  p-value: < 2.2e-16
```

Fig. 10 - Sortie textuelle modèle Tour2

```
Call:
lm(formula = Index ~ Tour3 * Série, data = data4)

Residuals:
    Min       1Q   Median       3Q      Max
-9.350 -1.901  0.034  1.825  5.696

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.39310    6.64753   0.059 0.9529
Tour3         0.13267    0.07373   1.799 0.0746 .
Série1       -22.78987    7.71296  -2.955 0.0038 **
Tour3:Série1   0.19013    0.08722   2.180 0.0313 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.559 on 115 degrees of freedom
Multiple R-squared:  0.7294,    Adjusted R-squared:  0.7224
F-statistic: 103.3 on 3 and 115 DF,  p-value: < 2.2e-16
```

Fig. 11 - Sortie textuelle modèle Tour3

Modèle de régression pour le Tour 1

Le modèle estimé pour le Tour1 est donné par :

$$Serie1 : \hat{m}(\text{Tour1}) = -21.82 + 0.33 \times \text{Tour1} \quad (5)$$

$$Serie2 : \hat{m}(\text{Tour1}) = 0.62 + 0.13 \times \text{Tour1} \quad (6)$$

Ces résultats montrent qu'une augmentation d'une unité du score au Tour 1 entraîne une hausse de 0,33 unités de l'index pour les joueurs de la Série 1 et de 0,13 unités pour ceux de la Série 2.

Modèle de régression pour le Tour 2

Le modèle estimé pour le Tour2 est donné par :

$$Serie1 : \hat{m}(\text{Tour2}) = -25.92 + 0.37 \times \text{Tour2} \quad (7)$$

$$Serie2 : \hat{m}(\text{Tour2}) = 1.62 + 0.11 \times \text{Tour2} \quad (8)$$

On observe que pour chaque unité de score supplémentaire au Tour2, l'index augmente de 0,37 unités pour la Série 1 et de 0,11 unités pour la Série 2. L'effet du Tour 2 sur l'index semble donc plus marqué que celui du Tour1.

Modèle de régression pour le Tour 3

Le modèle estimé pour le Tour3 est donné par :

$$Serie1 : \hat{m}(\text{Tour3}) = -22.40 + 0.32 \times \text{Tour3} \quad (9)$$

$$Serie2 : \hat{m}(\text{Tour3}) = 0.39 + 0.13 \times \text{Tour3} \quad (10)$$

On constate qu'une augmentation d'une unité du score au Tour3 entraîne une hausse de 0,32 unités de l'index pour la Série 1 et de 0,13 unités pour la Série 2.

Conclusion sur la modélisation

On observe que les scores obtenus aux Tours 2 et 3 ont un impact plus important sur l'index que ceux du Tour 1. De plus, l'effet du score diffère selon la Série des joueurs, ce qui justifie l'introduction d'une interaction dans notre modèle.

3 - Analyse des variations de l'index en fonction des résultats obtenus lors des trois tours

3.1 Modèle de régression linéaire multiple

Nous proposons un modèle de régression linéaire multiple pour estimer l'index d'un joueur en fonction des résultats obtenus lors des trois tours. La relation entre ces variables est modélisée par l'équation suivante :

$$m(\text{Tour1}, \text{Tour2}, \text{Tour3}) = \beta_0 + \beta_1 \times \text{Tour1} + \beta_2 \times \text{Tour2} + \beta_3 \times \text{Tour3} + \varepsilon \quad (11)$$

L'ajustement du modèle sur les données nous donne l'estimation :

$$\hat{m}(\text{Tour1}, \text{Tour2}, \text{Tour3}) = -39.68 + 0.19 \times \text{Tour1} + 0.24 \times \text{Tour2} + 0.12 \times \text{Tour3} \quad (12)$$

3.2 Test de la significativité du modèle par rapport au modèle constant

Nous effectuons une analyse de variance (ANOVA) pour comparer notre modèle de régression multiple avec un modèle constant. Voici les 2 hypothèses du test :

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_1 : \exists i \in \{1, 2, 3\} : \beta_i \neq 0 \quad (13)$$

Avec une statistique de Fisher de 103,6 et une p-valeur extrêmement faible ($p < 2,2 \times 10^{-16}$), nous rejetons l'hypothèse nulle, ce qui signifie que notre modèle complet est significativement meilleur que le modèle constant.

3.3 Interprétation des coefficients et tests de Student

Voici les hypothèses pour le test de Student :

$$H_0 : \beta_i = 0 \quad \text{vs} \quad H_1 : \beta_i \neq 0 \quad (14)$$

Les tests de Student montrent que toutes les variables explicatives ont un effet significatif sur l'index ($p\text{-value} < 0,05$).

L'interprétation de la sortie est la suivante :

- Un point supplémentaire au Tour1 entraîne en moyenne une augmentation de 0,191 de l'index.
- Un point supplémentaire au Tour2 entraîne en moyenne une augmentation de 0,239 de l'index.
- Un point supplémentaire au Tour3 entraîne en moyenne une augmentation de 0,116 de l'index, ce qui est plus faible que les autres tours.
- $R^2 = 0,7299$, cela signifie que le modèle explique environ 72,99% de la variabilité de l'index des joueurs. Le modèle est donc relativement performant pour expliquer les variations de l'index.

L'analyse montre que les résultats obtenus aux trois tours influencent significativement l'index des joueurs. Le modèle de régression multiple est bien ajusté et explique une grande part de la variabilité de l'index ($R^2 = 0,7299$). Tous les coefficients sont statistiquement significatifs, bien que l'impact du troisième tour soit plus faible par rapport aux deux premiers.

3.4 Visualisation des effets avec les diagrammes de régression partiels

Les trois graphiques ci-dessous (cf. Figure 12) permettent d'analyser l'influence des variables Tour1, Tour2 et Tour3 sur la variable dépendante Index, tout en neutralisant l'effet des autres

variables du modèle. Chaque graphique représente une régression partielle, mettant en évidence la relation entre l'index et une variable explicative spécifique après ajustement des autres effets.

- **Premier graphique (Index vs Tour1 | Tour2, Tour3)** : La relation semble linéaire et affiche une tendance positive, traduisant une corrélation croissante entre Index et Tour1. Le (R^2) partiel est de **15,3%**, indiquant une contribution modérée de Tour1 à l'explication de la variabilité de l'index.
- **Deuxième graphique (Index vs Tour2 | Tour1, Tour3)** : La régression linéaire est bien marquée et croissante, mettant en évidence un lien plus fort entre Index et Tour2. De plus, le R^2 partiel atteint **22%**, suggérant que Tour2 a une influence plus significative sur l'index.
- **Troisième graphique (Index vs Tour3 | Tour1, Tour2)** : Bien que la régression linéaire soit également croissante, son R^2 partiel reste faible (**5,1%**), ce qui indique une contribution mineure de Tour3 à l'explication de la variabilité de l'index.

En résumé, Tour2 semble être le facteur le plus influent, suivi de Tour1, tandis que Tour3 a un impact plus limité.

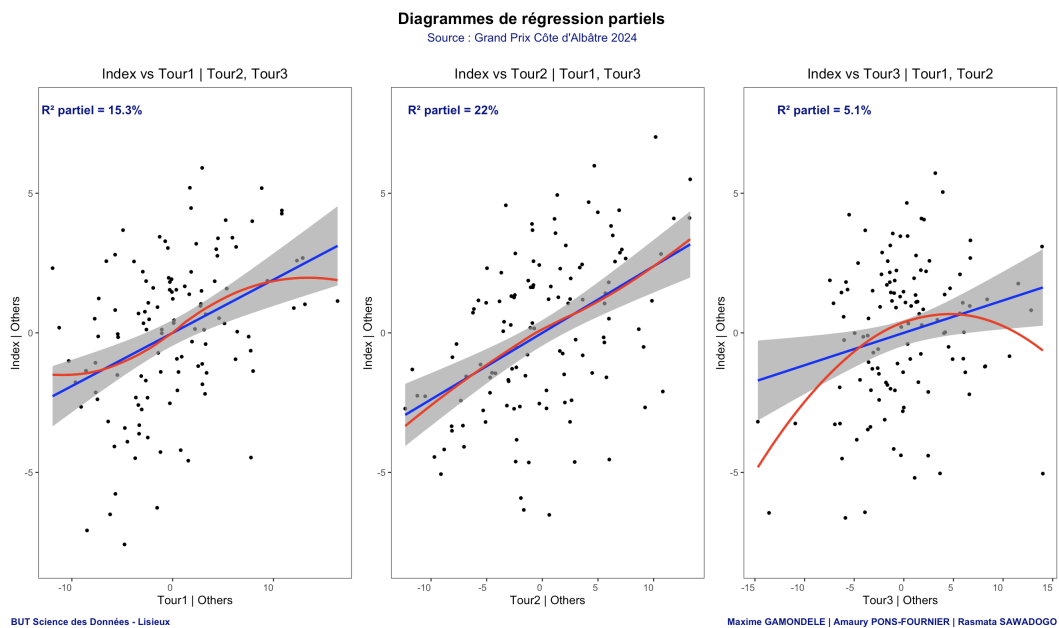


Fig. 12 - Diagrammes de régressions partiels

Conclusion

L'analyse des résultats du Grand Prix de la Côte d'Albâtre a permis de mettre en évidence plusieurs tendances intéressantes concernant la performance des joueurs. L'étude du lien entre le club d'origine et les scores obtenus a montré qu'aucune différence significative n'existait entre les clubs les plus représentés. Les résultats observés sont donc principalement attribuables aux performances individuelles plutôt qu'à une influence liée à l'appartenance à un club particulier.

L'exploration de la relation entre l'index des joueurs et leurs scores a mis en avant une corrélation positive. Plus l'index d'un joueur est élevé, plus son score total a tendance à être élevé, ce qui confirme que cet indicateur est un facteur clé pour anticiper les performances en compétition.

Enfin, l'étude de la relation entre les résultats du premier et du troisième tour a révélé une certaine stabilité des performances des joueurs tout au long du tournoi. Cette tendance souligne que les joueurs maintiennent un niveau de jeu relativement constant d'un tour à l'autre, malgré les variations pouvant être liées aux conditions de jeu ou à la pression compétitive.

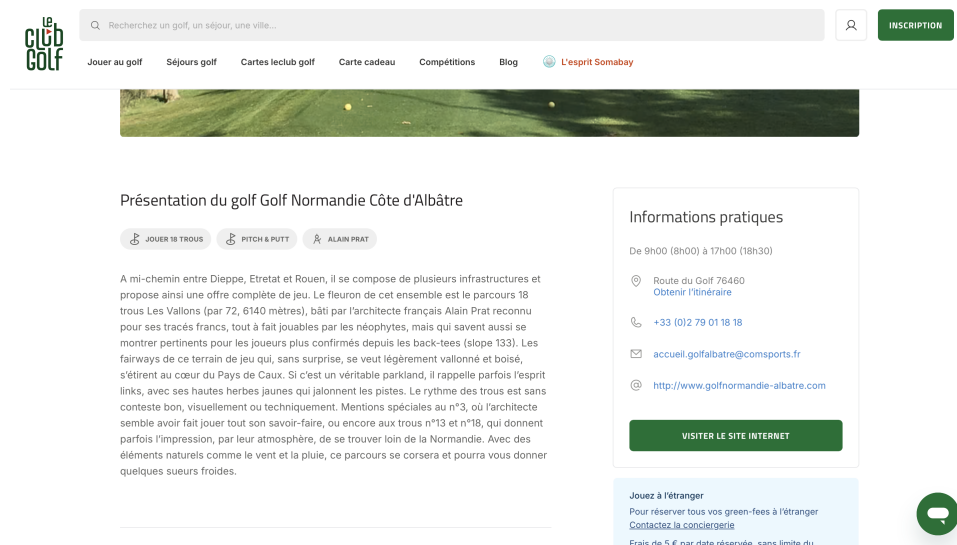
Ces analyses apportent des éléments de compréhension précieux pour les organisateurs du Grand Prix ainsi que pour les joueurs eux-mêmes. Elles pourraient être approfondies en intégrant d'autres facteurs, comme les conditions météorologiques ou les caractéristiques des parcours, afin d'obtenir une vision plus complète des éléments influençant la performance en compétition.



Annexe

Le Golf de Normandie Côte d'Albâtre est un parcours situé sur la côte normande, apprécié pour sa vue panoramique et son cadre naturel exceptionnel. Ce golf est également mis en avant sur le site Le Club Golf, qui répertorie les golfs de France et fournit des informations sur les tarifs, les réservations, ainsi que les services disponibles pour les golfeurs.

La capture d'écran ci-dessous provient du site **Le Club Golf** et présente les green fees (tarifs d'accès au parcours) du Golf de Normandie Côte d'Albâtre. Elle met en évidence les différentes options tarifaires selon les périodes de l'année et les types de prestations proposées.



The screenshot shows the website 'Le Club Golf' with a search bar at the top. Below the navigation menu, there is a large image of a golf course. The main content area is titled 'Présentation du golf Golf Normandie Côte d'Albâtre'. It features three tabs: 'JOUER 18 TROUS', 'PITCH & PUTT', and 'ALAIN PRAT'. The 'JOUER 18 TROUS' tab is selected. The text describes the golf course, mentioning its location between Dieppe, Etretat, and Rouen, and its design by Alain Prat. It highlights the course's features, such as its 18 holes, the 'Les Vallons' par 72, and the 'back-tees' (slope 133). The text also mentions the course's location in the Pays de Caux, its beautiful views, and its facilities, including a clubhouse and a restaurant. The text concludes by stating that the course is a must-visit for golfers in the region.

Présentation du golf Golf Normandie Côte d'Albâtre

JOUER 18 TROUS | PITCH & PUTT | ALAIN PRAT

A mi-chemin entre Dieppe, Etretat et Rouen, il se compose de plusieurs infrastructures et propose ainsi une offre complète de jeu. Le fleuron de cet ensemble est le parcours 18 trous Les Vallons (par 72, 6140 mètres), bâti par l'architecte français Alain Prat reconnu pour ses tracés francs, tout à fait jouables par les néophytes, mais qui savent aussi se montrer pertinents pour les joueurs plus confirmés depuis les back-tees (slope 133). Les fairways de ce terrain de jeu qui, sans surprise, se veut légèrement vallonné et boisé, s'étirent au cœur du Pays de Caux. Si c'est un véritable parkland, il rappelle parfois l'esprit links, avec ses hautes herbes jaunes qui jalonnent les pistes. Le rythme des trous est sans conteste bon, visuellement ou techniquement. Mentions spéciales au n°3, où l'architecte semble avoir fait jouer tout son savoir-faire, ou encore aux trous n°13 et n°18, qui donnent parfois l'impression, par leur atmosphère, de se trouver loin de la Normandie. Avec des éléments naturels comme le vent et la pluie, ce parcours se corsera et pourra vous donner quelques sueurs froides.

Informations pratiques

De 9h00 (8h00) à 17h00 (18h30)

Route du Golf 76460
[Obtenir l'itinéraire](#)

+33 (0)2 79 01 18 18

accueil.golfaibatre@comsports.fr

<http://www.golfnormandie-albatre.com>

VISITER LE SITE INTERNET

Jouez à l'étranger
Pour réserver tous vos green-fees à l'étranger
[Contactez la conciergerie](#)
Prix de 5 € par date réservée, sans limite de

Table des figures

1	Analyse des index des joueurs des cinq clubs les plus représentés	5
2	Relation entre le score total et l'index	6
3	Relation entre les résultats du tour 3 et du tour 1	7
4	Relation entre les résultats du tour 3 et du tour 1	8
5	Relation entre les résultats du tour 3 et du tour 1	9
6	Relation entre les résultats du tour 3 et du tour 1	10
7	Extrait du jeu de données	11
8	Relations entre l'index et les scores des trois tours	12
9	Sortie textuelle modèle Tour1	13
10	Sortie textuelle modèle Tour2	13
11	Sortie textuelle modèle Tour3	13
12	Diagrammes de régressions partiels	16