

J'ai donc choisi d'étudier le Norseman - triathlète oblige autant ne pas prendre le plus facile

Il se trouve que Athlinks n'a pas les résultats de toutes les éditions. Il a donc fallu trouver un autre moyen d'avoir accès à ces résultats. Il se trouve que l'organisation du Norseman rend publique l'ensemble des résultats de toutes les années précédentes dans un même fichier excel téléchargeable.

Pour mon étude de cas, je me suis plongé dans les facteurs de performance de la course menant au *black t-shirt*. Mais certaines années, en raison de la météo menaçante en haut du Gaustatoppen (orage le plus souvent) l'accès y est interdit, et tous les concurrents finissent au même endroit plus bas, ligne habituellement réservée aux *white t shirt finishers*. J'ai donc choisi de ne pas prendre en compte ces éditions (2005, 2007 et 2023)

Cette base de données ne contient pas non plus les données météo. Etant donné que j'ai étudié plusieurs années de cours et que les conditions sont différentes selon toutes les années, il aurait été difficile de prendre en compte cette donnée dans la performance.

Cependant, comme on parle surtout ici de classement par édition et non de temps applicable à l'ensemble des éditions, il apparaît que la météo devient "négligeable" dans le sens où elle est la même pour tous les concurrents de cette même édition. Bien sûr il y a sûrement des bons ou des mauvais réponds aux conditions.

Des récits de course que j'ai lu, il n'y a eu que peu d'éditions avec des conditions extrêmes et celles-ci n'ont pas été prises en compte.

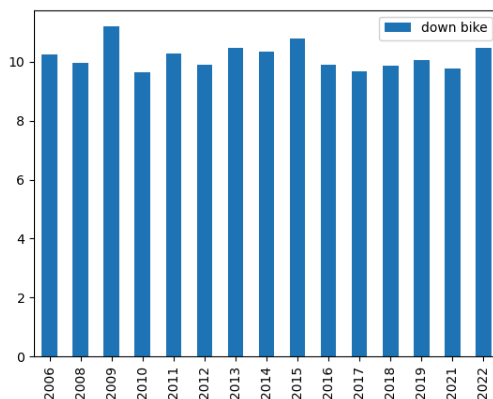
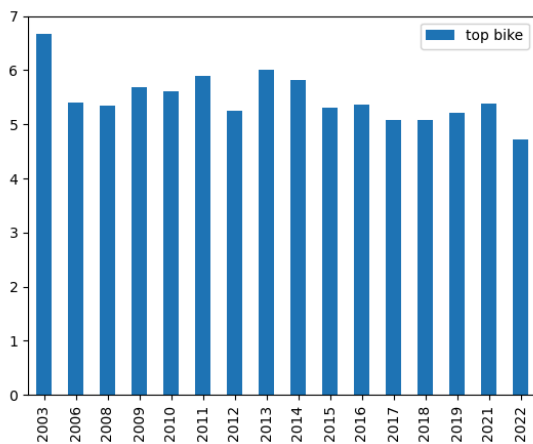
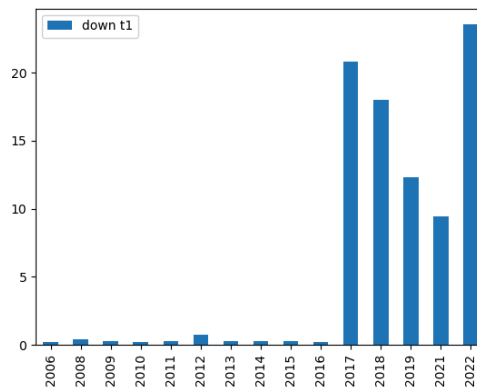
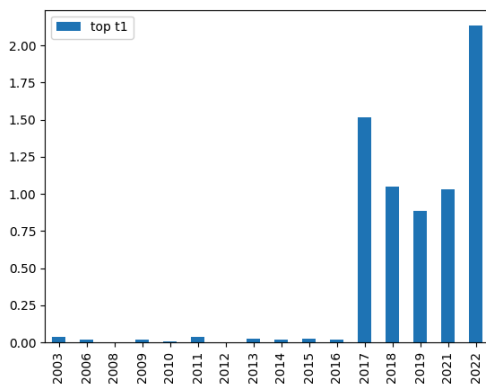
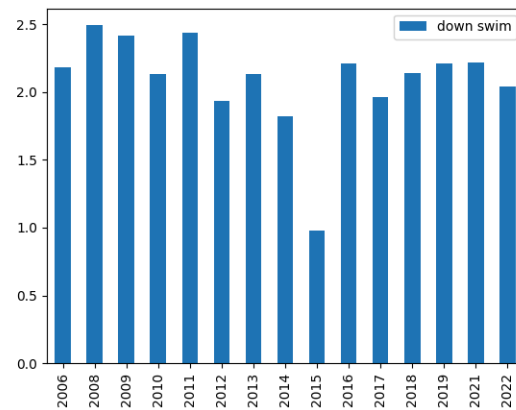
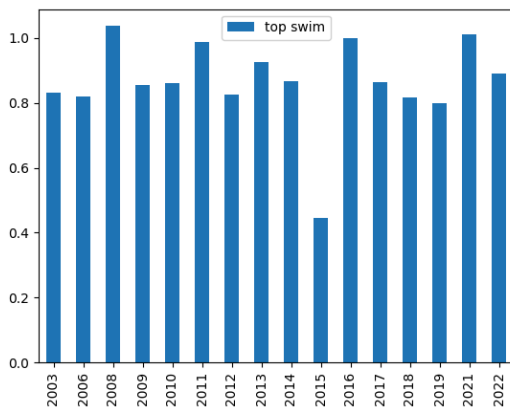
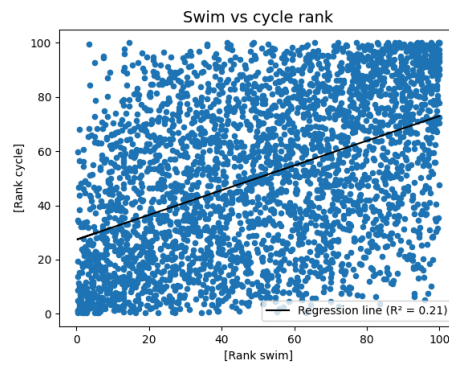
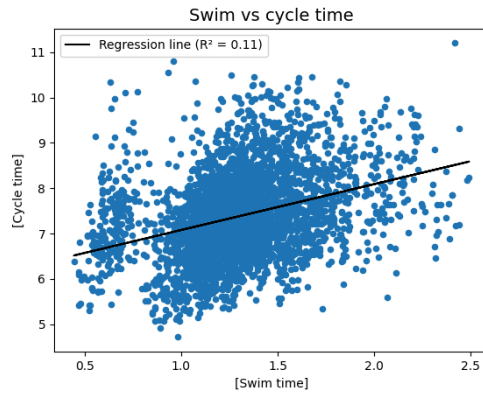
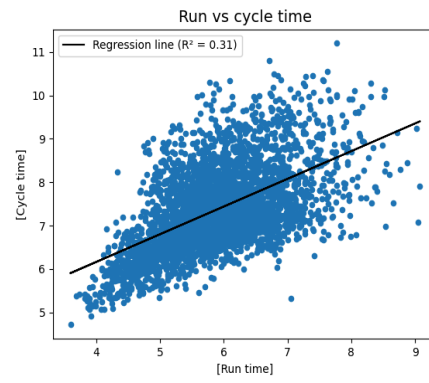
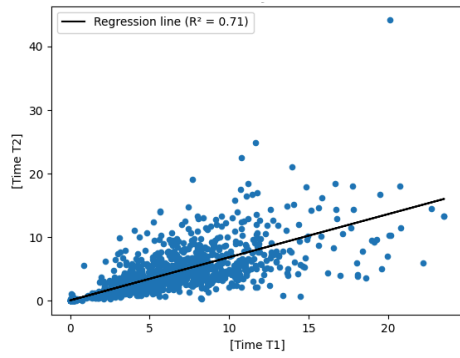
Je n'ai pas non plus pris en compte les temps de transition puisque ceux-ci sont erronés en partie : pour une partie des athlètes, les heures ont été prises en compte pour des minutes, ce qui fausse les temps intermédiaires. Peu de différences se font sur une transition sur une course aussi longue.

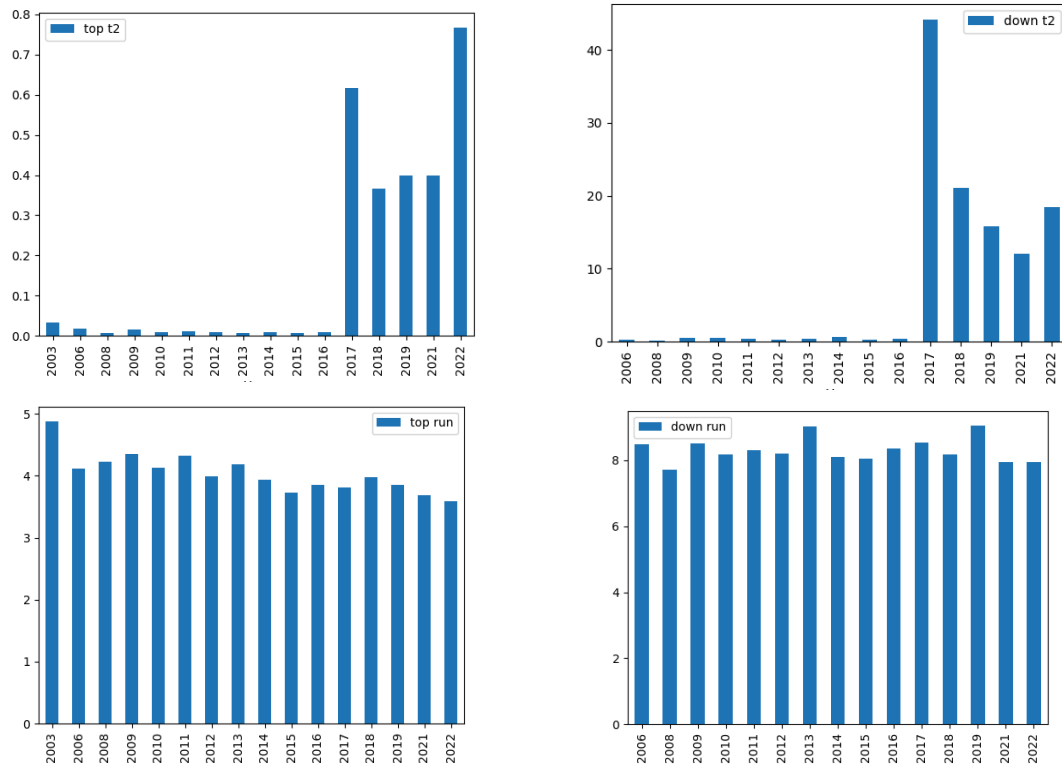
J'aurai aimé avoir accès aux temps intermédiaires de l'épreuve pour affiner l'analyse (e.g. temps en haut de cols à vélo, temps de la partie plate en course vs partie montante).

Concernant l'analyse, j'ai d'abord calculé les temps cumulés après le vélo puis fait les classements par année de chaque concurrent pour chaque discipline, ramené à un pourcentage du total de l'année pour pouvoir comparer les différentes années (il y a eu 2 éditions en 2012 par exemple donc 400 concurrents).

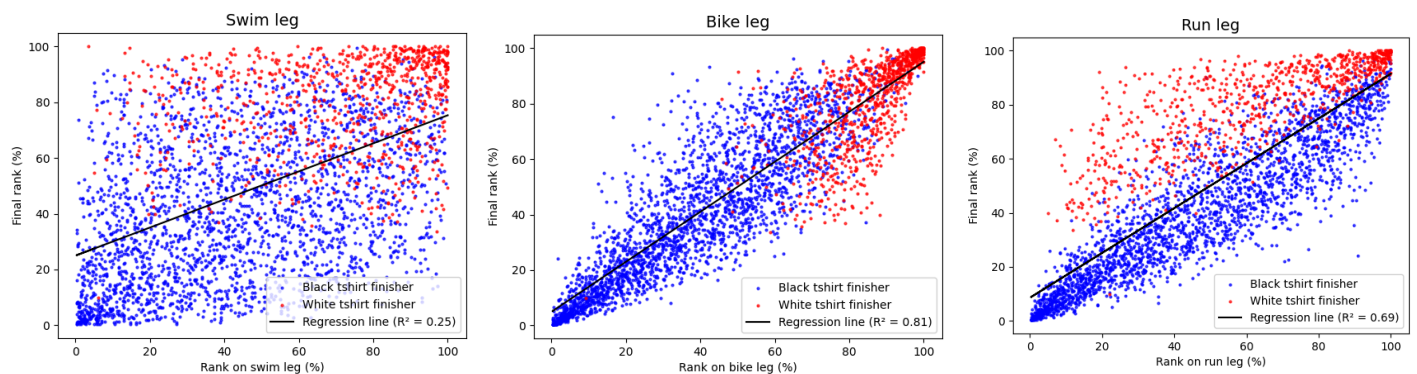
La phase exploratoire a consisté à croiser les différents temps et rangs des disciplines pour en observer les corrélations, ainsi que voir les temps minimum et maximum de l'épreuve chaque année, voir si les variations étaient importantes (graph ci-dessous).

Une fois cette phase faite, elle a permis de voir les erreurs de chrono sur les transitions et que ces erreurs sont aléatoires, d'où le choix de ne pas prendre en compte ces chronos (on ne sait pas si il a vraiment passé 1h en transition ou pas)





Il en est ressorti que la partie vélo était la plus importante, la partie incontournable pour performer comparé aux parties natation et course, même si cette dernière reste importante on reste plus dans la gestion de la position que dans la possibilité de gagner beaucoup de places

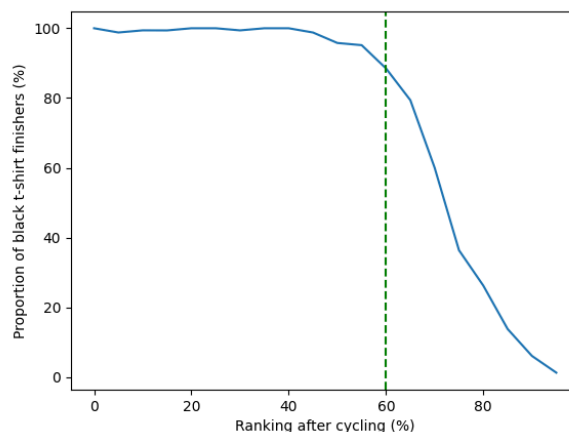


J'ai choisi une régression linéaire pour mesurer la corrélation entre les variables dépendantes 2 par 2.

J'aurais sûrement opté pour un modèle plus complexe si j'avais eu en plus les données intermédiaires dans chaque discipline.

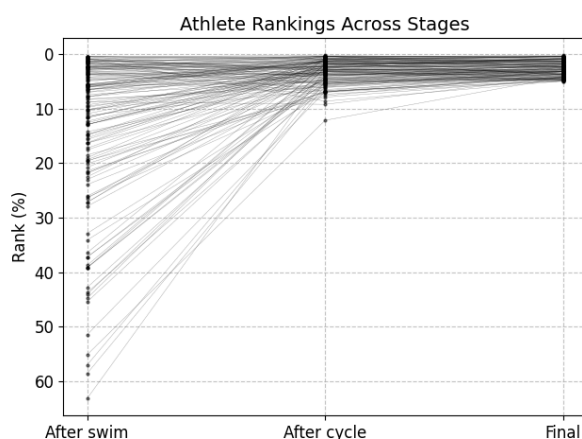
Et cette corrélation à l'avantage d'être très visuelle dans les cas où les corrélations sont bonnes comme ici et donc plus faciles à intégrer dans un article pour expliquer ou montrer la corrélation facilement à tous les publics.

Une fois cette recherche de paramètres terminée, je suis passé aux recommandations pratiques avec les chances d'atteindre le top 160 en fonction du rang à la fin du vélo et donc avoir un objectif de temps à avoir à la fin de cette partie et par déduction avoir une idée des temps qu'il faut pour chaque discipline. au vu du peu de variation des temps mis par les derniers black finishers, on peut se dire que ces prédictions sont plutôt bonnes.



Si l'on prend un seuil des 60 % qui semble être un bon seuil pour garder de bonnes chances de se retrouver en bonne position au cut-off, et en analysant les données des athlètes de ce rang des années précédentes : natation et T1 en 1h15 ; vélo en 7h30 (24 km h) ; course en 6h10. Même si bien sur ces prédictions peuvent changer en fonction du courant en natation, du vent en vélo... etc

Enfin j'ai voulu m'intéresser au top 5%. En étant surpris de l'écart de niveau en natation et on retrouve bien que la remontée en vélo dans ces cas là est très importante. une telle remontée est impossible en course à pied à ce niveau puisque tous sont aussi de bons coureurs (seuls 4% hors du top15 après le vélo ont fini dans le top10)



C'est une étude qui manque peut être de précision au vu des seules données de départ et d'arrivée et du manque d'informations contextuelles. Mais elle donne une bonne idée des performances nécessaires et des qualités requises pour ramener un tshirt noir de cette épreuve.