

Julien Latraverse
LATJ07119500

MAKING-OF

Pourquoi ce sujet?

Il existe dans le monde du hockey plusieurs mythes entourant le repêchage des joueurs de la ligue nationale (LNH). Par exemple, un de ces mythes affirme que le dépisteur du Canadien, Trevor Timmins, fait des choix de piètre qualité au repêchage. Un autre affirme qu'une équipe, comme les Penguins de Pittsburgh ou les Capitals de Washington, reconnue pour la qualité de son "draft" se voit assurée de remporter la Coupe Stanley. Mais est-ce vraiment le cas? Notre reportage a ainsi pour but de confirmer si ces légendes urbaines ont lieu d'exister ou si elles ne sont que des commérages de gérants d'estrade.

Quels outils ou technologies avez-vous utilisées?

Pour arriver à nos fins, il fallait recenser l'ensemble des données des joueurs repêchés. Cependant, un joueur repêché en 1982 ne démontre pas réellement la réalité du hockey moderne. Nous avons donc décidé de commencer notre data-scraping à partir de 2004. Pourquoi 2004? À cause du [lock-out](#) qui marque le début d'une sorte de nouvelle ère du hockey moderne tel que nous le connaissons aujourd'hui.

Au début de notre projet, nous voulions utiliser un [API](#) de la LNH pour effectuer notre moissonnage de données. Celui-ci manquait malheureusement une information cruciale pour nos calculs: le nombre de matchs joués par joueur.

Cette donnée permet de recenser si un joueur a joué au moins une saison dans la ligue. En effet, un des barèmes utilisé dans le reportage pour identifier l'impact d'un joueur dans une équipe est d'établir sa moyenne de points par match. On parvient alors à relativiser les données entre les

vétérans, qui ont joué plus de matchs, les jeunes. De plus, on évite toutes anomalies, comme la moyenne de trois buts par match après une partie de [Ryan Poehling](#).

C'est le site de statistiques [hockey-reference](#) qui nous fournira toutes les données nécessaires pour notre data-scraping. À l'aide des précieuses [notes de cours](#), de la patience de notre tuteur et de BeautifulSoup nous avons finalement réussi à écrire notre script Python.

Grâce à lui, nous avons pu inscrire toutes nos données dans un fichier .csv (*nommé en l'honneur d'un de mes joueur préféré, [Mark Scheifele](#)*). À ce moment, la partie la plus ardue était derrière nous. Il ne restait plus qu'à interpréter cette immense quantité de données à l'aide d'un carnet Jupyter. Nous avons préféré [Jupyter](#) pour l'aisance avec laquelle il était possible de filtrer les données. En une seule ligne de code, on pouvait faire ressortir tous les premiers choix au repêchage ou isoler tous les joueurs ayant joué plus de 82 matchs et plus encore.

De ce fait, [Jupyter](#), nous a aussi permis en quelques lignes - *que dieu bénisse la fonction groupby!*- de voir que le Canadien de Montréal se retrouve souvent dans les quinze meilleures équipes au repêchage ou de comprendre pourquoi les équipes comme les Blackhawks, les Kings ou les Penguins ont remporté plusieurs [Coupes Stanley depuis 2004](#).

Obstacles

Le plus grand obstacle rencontré pour notre reportage est la compréhension de Python et Jupyter de Thomas et moi. Nous n'avions aucune expérience préalable dans ce type de travail et de langage. D'où la nécessité pour nous de choisir un sujet simple, mais intéressant. On avançait souvent à tâton et à grands coups d'essais et d'erreurs, certes, mais on a tout de même réussi à écrire le script. Après cette barrière, l'interprétation de nos données était facilement réalisable et n'a nécessité que quelques conseils d'appoints.