UNIVERSITÉ DU QUÉBEC À MONTRÉAL

*Making-of*

TRAVAIL PRÉSENTÉ À

JEAN-HUGUES ROY

DANS LE CADRE DU COURS

Technologies de l'information appliquées au journalisme

EDM5240

GROUPE 20

PAR

THOMAS PICOTTE-LAVOIE (PICT14099508)

17/04/2019

**Le repêchage : une science inexacte ?**

Le monde du sport professionnel est de plus en plus influencé par l’analyse des statistiques avancées et du journalisme de données. On tente de calculer avec le plus d’exactitude possible la performance et le rendement d’un joueur ou d’une équipe. Ce désir grandissant de quantification des données s’inscrit dans une tendance assez évidente, celle de la parité. La compétition est si féroce qu’on doit se démarquer par tous les moyens. Chaque détail compte. Le monde du hockey n’y fait pas exception.

Aux quatre coins de la Ligue nationale de hockey (LNH) et de la sphère médiatique qui l’entoure, il existe deux écoles de pensée : la plus récente, qui croit ardemment aux statistiques avancées et à l’existence d’un lien entre les chiffres relatifs à la possession de rondelle et les succès. Puis, il y a l’ancienne école de pensée, la « vieille école », qui aborde les statistiques avancées avec un certain scepticisme. L’ardeur au travail et l’intensité le long des rampes, ça ne se calculerait pas. Le talent, la vitesse et la robustesse remportent des championnats, pas le pourcentage de tirs tentés.

Qui a raison ? Qui a tort ?

C’est dans cet esprit et avec cette question en tête que Julien et moi avons décidé d’aborder un aspect précis du sport national des Canadiens qui contribue plus souvent qu’autrement au succès d’une équipe : le repêchage amateur.

**Pourquoi ce sujet ?**

On entend souvent que le repêchage est une « science inexacte ». Cette expression largement galvaudée a-t-elle sa raison d’être ?

Alors que la direction du club de hockey Canadiens amorce un virage jeunesse et une phase de reconstruction, on s’est demandé si ses succès au repêchage pourraient contribuer à lui redonner ses lettres de noblesses dans l’espoir d’une atteinte des gloires passées.

Nous nous sommes donc posés les questions suivantes :

1. Quelle équipe repêche le mieux depuis 15 ans ?
2. Où se situe le Tricolore comparativement à ses homologues de la LNH ?
3. Quel pays / ligue / équipe junior produit le plus d’espoirs ?
4. Comment l’analyse des statistiques avancées influence les choix au repêchage et le sport en général (du point de vue d’un recruteur et d’un journaliste sportif) ?
5. À quel point le repêchage permet de remporter des championnats ?

**Outils utilisés**

Nous avons d’abord effectué quelques recherches pour trouver un API ou une base de données comprenant l’entièreté des choix au repêchage de la Ligue nationale de hockey (LNH).

Nous avons rapidement constaté que les données étaient limitées, incomplètes ou simplement inexistantes. On a donc opté pour l’utilisation de la bibliothèque python *BeautifulSoup*, puisqu’elle nous permettait d’extraire des informations en provenance du site web *Hockey Reference* (<https://www.hockey-reference.com/draft/>), une source fiable et respectée dans le domaine des statistiques sportives.

On y retrouve la totalité des choix au repêchage de la LNH depuis 1963, mais nous avons décidé de limiter nos recherches à un spectre moins volumineux en étudiant le repêchage sur une période de 15 ans, de 2004 à 2018.

Pourquoi cet échantillon, précisément ? Pour deux raisons précises :

1. Julien et moi avons commencé à suivre le hockey de la LNH après le lock-out de 2004-2005, cet arrêt de travail désormais célèbre qui avait privé les amateurs de leur pain quotidien pendant une saison complète. Ce point de départ commun nous semblait un choix logique.
2. Le hockey a énormément changé en 15 ans. Une nouvelle génération de joueurs a éclos et le jeu lui-même a évolué drastiquement.

**Problèmes éprouvés**

On a longtemps cherché à travailler à partir d’un API, avant de réaliser que l’option *BeautifulSoup* était une alternative parfaite à notre projet. Le contexte particulier de cette session d’hiver – vive la grève ! – nous aura grugé du temps et des notions précieuses qui auraient grandement contribué à l’efficacité de nos projets.

Vu nos connaissances limitées, nous avons préféré opter pour des bases simples mais concrètes afin de s’amuser avec les données et ainsi faire ressortir plusieurs informations pertinentes (le pays qui produit le plus de joueurs repêchés, les cohortes les plus prolifiques, les équipes ayant produit le plus de joueurs, etc.)

Nous voulions éviter de viser trop haut pour finalement se noyer dans une marre de données qu’on ne comprenait pas. Surtout que nous ne pouvions compter sur l’expertise des étudiants de Polytechnique.

Nous avons connu quelques pépins techniques quand est venu le temps d’ajouter des variables dans un carnet Jupyter, mais avec l’aide de certains camarades de classe et de périodes de tutorat, les problèmes encourus ont été facilement réglables.