



Maxime Hebert

Making-of

Mon projet de fin de session est un peu la continuité de ce que j'ai entrepris lors du moissonnage de mi-session. À la mi-session, j'avais moissonné tous les salaires des joueurs de la Ligue nationale de hockey (LNH) depuis 2008.

Pour pousser plus loin ce que j'avais fait en milieu de session, j'ai donc choisi de trouver les joueurs les plus rentables dans la Ligue. Dans une ère de plafond salarial, chaque signature a des conséquences. Les directeurs généraux tentent donc chaque année de trouver des merveilles à bas prix pour demeurer sous le plafond salarial. En ce sens, je voulais ainsi savoir quelles étaient les meilleures signatures des DG.

Pour débiter mon projet, j'ai utilisé Cloud 9, et donc Python, afin de moissonner toutes les données nécessaires. D'abord les salaires des joueurs (cette fois juste pour 2017), puis les statistiques de ces derniers.

J'ai ensuite mis tous les joueurs dans Excel (Numbers sur Mac) et j'ai jumelé les statistiques et les salaires de chaque joueur. J'ai choisi Excel pour continuer le projet, car avec ce programme, les calculs se font très facilement et au niveau visuel, c'est le meilleur programme.

Avec toutes les données assemblées, j'ai pu commencer finalement mes calculs. Pour chaque joueur, j'ai divisé le salaire par le nombre de points des joueurs afin d'obtenir le coût par point. Ensuite, j'ai divisé le nombre de points par le nombre de matchs disputés par les joueurs pour savoir le nombre de points par match. Au final, sans surprise, c'est l'attaquant Connor McDavid qui a été le joueur le plus rentable de la LNH avec un coût de 9250\$ par point et 1,22 point par joute.

J'ai cependant voulu aller plus loin que les joueurs individuellement. J'ai tenté de savoir quelle équipe avait la formation la plus rentable globalement. J'ai d'abord utilisé trois listes. La première constitue tous les joueurs de la LNH, la deuxième, tous ceux qui ont joué au moins 20 matchs (le quart de la saison), et la troisième, tous ceux qui en ont disputé au moins 41 (la moitié de la saison). Cela m'a permis d'éliminer les joueurs qui n'ont été rappelés par leur formation que le temps de quelques matchs ainsi que les blessures majeures.

J'ai par la suite créé un nouveau tableau où j'inscrivais le coût par point moyen des 30 équipes, que j'ai classé de 1 à 30. J'ai fait le tout pour chacune des listes décrites précédemment pour finalement faire un classement global des équipes. J'ai ajouté leur classement général dans la Ligue nationale pour voir si selon leur position dans celui-ci, si ça influençait leur position au niveau de la rentabilité.

Comme j'ai vu qu'individuellement, ce sont les jeunes qui sont les plus rentables, j'ai ajouté la moyenne d'âge de chacune des équipes afin de voir si cela avait un réel impact sur le classement. À ma grande surprise, c'est même l'inverse : ce sont pour la grande majorité des équipes avec une haute moyenne d'âge qui sont les plus rentables. Comme quoi l'expérience joue un grand rôle sur le salaire et que celui-ci est justifié.

En terminant, j'ai fait un autre tableau avec le nombre de points et la moyenne de point par match de chacune des équipes. Je voulais grâce à ce tableau voir le lien entre la productivité et la rentabilité.

J'ai rencontré deux problèmes durant mon projet. Le premier a été de moissonner les données statistiques des joueurs de la LNH. En effet, les sites tels que NHL.com, RDS.ca et sportsnet.ca n'offrent pas de *pattern* dans leur URL me permettant de réussir à moissonner facilement les données. Il m'a fallu quelques recherches avant de trouver un forum indiquant que le site hockey-reference.com offrait une version CSV des statistiques. Le deuxième problème que j'ai rencontré durant la production de mon travail final a été plus personnel. Voyant que mes trucs fonctionnaient, j'ai toujours cherché à aller plus loin dans mes calculs plutôt que de me concentrer sur une ligne directrice précise. J'en suis venu à me poser plus de questions qu'autre chose au final. Cela a d'ailleurs fait en sorte que j'ai manqué de temps pour créer un reportage sur le sujet.

Malgré tout, à la lumière de ce que j'ai accompli avec mes données, je suis très satisfait de mon travail et je sais qu'au moins, il y a quelque chose à faire avec ça !

