

MAKIN-OF

Dans le cadre de ce travail final, j'ai décidé de faire l'analyse des statistiques des cinq plus grandes ligues européennes de soccer. Celles-ci me permettront de comparer le style de jeu prisés par les clubs de chacun de ces circuits.

Cette idée m'est venue du fait que les clubs espagnols dominent les grandes compétitions européennes depuis quelques saisons. En effet, la Ligue des champions de l'UEFA et la Ligue Europa ont respectivement été remportées à sept et six reprises par des équipes provenant de ce pays au cours des dix dernières éditions. Je souhaitais donc analyser le style de jeu espagnol à l'aide de statistiques et voir si celui-ci se reflétait dans les résultats européens.

Pour faire le tout, j'ai déniché des fichiers CSV de la [Liga](#), la [Premier League](#), la [Bundesliga](#), la [Ligue 1](#) et la [Serie A](#). Les fichiers de la Ligue des champions de l'UEFA étaient seulement disponibles sous format payant, donc j'ai dû dénicher les statistiques sous diverses plateformes, notamment pour [avoir accès à la moyenne de buts marqués](#) et [aux finalistes](#) de chacune des éditions de la compétition.

J'ai ensuite utilisé Visual Studio Code pour faire l'analyse des différentes statistiques. Initialement, j'ai incorporé les 50 fichiers dans le logiciel pour faire le traitement des données. Cela m'a permis de trouver le nombre de victoires à domicile, à l'étranger et les matchs nuls de chacune des ligues sur les dix dernières saisons, tout en y ajoutant le pourcentage. Cela a toutefois été un exercice plutôt long, car j'utilisais une boucle pour chacun des fichiers. Cela équivalait à plus de 700 lignes de code pour une seule information parmi celles que je souhaitais trouver.

La période de tutorat a donc été bien utile. À la suite de celle-ci et des tutoriels reçus, j'ai pu cumuler les autres statistiques bien plus rapidement. Tous les fichiers

étaient désormais regroupés sous un même fichier (futbol.csv) et des formules qui n'avaient pas été vues en classe m'ont permis de compléter ma recherche. Celles-ci impliquaient notamment l'utilisation de l'extension Panda qui m'a permis de faire la somme des statistiques se retrouvant dans chacune des colonnes de mon CSV. Ces résultats étaient divisés de façon à obtenir les statistiques à domicile, à l'extérieur, le total et la moyenne par rencontre.

J'ai donc obtenu le cumulatif de chacune des ligues pour les résultats, les buts, les fautes et les cartons. J'ai aussi comptabilisé la moyenne des cotes de paris pour dénicher d'autres informations intéressantes. J'ai notamment remarqué que dans les ligues où les équipes à domicile gagnent davantage de matchs, soit la Liga et la Premier League, les cotes sont étrangement plus élevées que dans les autres championnats.

À la suite de l'analyse des données recueillies, j'ai fait un plan de mon texte et contacté des pistes potentielles d'intervenants. Nilton Jorge, journaliste sportif, m'a répondu et, à la suite de notre entrevue, j'ai pu bonifier mon article à l'aide de ses propos. J'ai aussi fait quelques montages visuels pour rendre le tout plus interactif.

Je suis satisfait du résultat final du travail, notamment car les statistiques dénichées permettent d'expliquer la domination espagnole dans le monde du soccer au cours des dix dernières années. De plus, aimant faire quelques paris de temps en temps, cela me permettra d'être mieux informé et, peut-être éventuellement, de créer un outil augmentant mon taux de réussite! ;-)