

Projet Analyse de données :

Présentation des données choisies :

Le set de données choisies pour ce projet est : l'Ultimate UFC Dataset.

A) Contexte du dataset :

l'UFC (Ultimate Fighting Championship) est la plus grande organisation de MMA du monde. Le MMA (Mixed Martial Arts, Arts Martiaux Mixtes en Français) est un sport de combat alliant différents sports de combat et arts martiaux, de percussion et de préhension (kickboxing, lutte, jiu jitsu brésilien...), se déroulant dans une cage octogonale, où, de plus, certaines frappes au sol sont autorisées. Un combat de MMA se déroule en 3 ou 5 rounds de 5 minutes, et peut être remporté de plusieurs manières différentes :

- par KO / TKO, lorsqu'un combattant n'est plus capable de se défendre intelligemment à la suite d'une ou plusieurs percussions (soit parce qu'il est inconscient, soit parce qu'il n'affiche plus aucune réponse défensive, même s'il n'est pas tout de suite inconscient).
- par soumission, lorsqu'un combattant abandonne à la suite d'une prise de soumission (étranglement, clé de bras, clé de jambe...), ou lorsque celle-ci entraîne perte de conscience ou blessure.
- par décision unanime/partagée. Lorsque la durée du combat est écoulée, un gagnant est nommé par les juges avec un système similaire à celui de la boxe (on donne un nombre de points aux combattants à chaque round, le gagnant est celui qui a le plus de points à la fin des rounds. Souvent, un juge donne 10 points au "gagnant" du round et 9 points au "perdant", Le perdant peut parfois se voir attribuer moins de points si il y a eu une grosse domination lors du round) Le résultat d'un combat est décidé par 3 juges, le gagnant gagne par décision unanime si les 3 juges l'ont désigné gagnant, et par décision partagée si un l'a désigné perdant

Ainsi, l'UFC organise des évènements de MMA dans le monde entier, et dans 13 catégories différentes (8 pour les hommes et 5 pour les femmes). Il existe ainsi pour chaque catégorie un classement des 15 meilleurs combattants de celle-ci ainsi qu'un champion. Les combats de l'UFC , comme pour beaucoup d'évènements sportifs, sont soumis à des paris, et il existe donc des cotes pour déterminer le vainqueur d'un combat mais aussi la manière ou les détails d'une victoire (par KO, au round 5 etc...).

B) Contenu du dataset.

Ce dataset (source : <https://www.kaggle.com/datasets/martj42/ufc-rankings>) est en fait une jointure de 3 datasets différents, l'un qui détaille l'organisation des combats et des combattants, l'autre qui détaille les cotes pour les paris, et la troisième détaille les classements potentielles des deux combattants d'un combat.

Ainsi, le dataset “UFC Ultimate Dataset” recense tous les combats organisés par l’UFC depuis l’année 2010 jusqu’à maintenant (la version qui sera utilisée pour ce projet voit sa dernière mise à jour prendre en compte les combats de la carte “UFC fight night : Moreno vs Albazi” du 2 Novembre 2024), à raison d’environ un évènement par semaine et un événement “numéroté” (qui correspond souvent à un événement où un ou plusieurs combats pour une ceinture se ont lieu) par mois. En tout, la base de données regroupe presque 6500 combats différents. Pour chaque combat, on a les détails de l’événement (ville, pays et date de l’événement), ainsi que sur le combat en lui-même :

le nom des combattants (qui sont appelés RedFighter et BlueFighter en fonction de leur “coin” de départ), la catégorie de poids du combat, le nombre de rounds, si le combat est un combat pour le titre, le résultat du combat, ainsi que la manière dont il s'est fini etc...

De plus, un grand nombre de colonnes sont dédiées à une étude plus approfondie des combattants en eux-même au moment du combat, comme par exemple leur série actuelle de victoire ou de défaite, la quantité moyenne de coups envoyés, touchés ou reçus, la quantité de takedowns (projection vers le sol), le nombre de victoire par soumission, la taille, l'allonge etc... le tout pour les deux combattants.

On a également des colonnes qui comparent directement certains aspects des deux combattants (qu'on pourrait obtenir en comparant directement les colonnes concernées pour les deux combattants), comme par exemple la différence d'allonge, en âge, ou en moyenne de takedown par combat. On a également plusieurs colonnes détaillant le rang d'un combattant dans chacune des catégories de l'organisation (il n'y a pas une seule colonne parce que certains combattants fluctuent entre deux catégories, comme par exemple Max Holloway actuellement classé 5^e chez les poids légers et 2^e chez les poids plume), et on a finalement différentes colonnes détaillant les cotes pour les différents résultats possibles du combat (victoire par décision, victoire par soumission etc...).

C) Question possibles pour ce dataset:

Ce dataset est utilisé au départ probablement pour les personnes cherchant à parier sur l'issue des combats, notamment sur un combat précis ou pour un combattant précis. On pourrait donc utiliser cette base de données pour essayer de justifier ou non les cotes données pour certains des combats recensés ou pour des futurs combats de l’UFC. Cependant il pourrait être utilisé pour d'autres choses, comme des questions plus générales. Par exemple :

- En considérant qu'un combat est plus amusant s'il se termine par un KO/TKO ou par une soumission, on peut comparer les combats par pays et par catégories de

poids afin d'essayer de trouver le moyen d'en avoir le plus pour son argent en choisissant quel évènement on veut voir.

- en MMA, il existe deux styles de combats très distincts, les 'strikers', des combattants qui vont chercher à rester debout et à utiliser différentes percussions pour gagner le combat par KO/TKO ou par décision, et les 'wrestlers' (lutteurs) ou 'grappler', qui vont plutôt chercher à amener leur adversaire au sol, souvent dans une position dominante, souvent pour finir le combat avec une soumission, des frappes au sol, ou une domination au sol. Ces deux styles sont très opposés, et certaines données permettraient avec cette base de données d'estimer si un style fonctionne mieux que l'autre et dans quelles conditions.