

Projet de fin de semestre

Description générale

Ce projet consiste à donner des outils d'analyse sur un data set portant les pronostics de professionnel concernant les courses de chevaux. Ce dataset contient chaque pronostic (quel cheval est censé gagner) et si ce pronostic s'est avéré juste ou non.

Provenance des données et méthodes d'obtention

Ce data set provient du site kaggle.com il est facilement téléchargeable après inscription via le liens :

<https://www.kaggle.com/gunner38/horseracing/version/1>

Structures attendues

- Charger de l'ensemble des données en une base de données mongoDB.
-
- Créer d'une base de données neo4j avec comme nœuds :
 - Le cheval,
 - Le pronostiqueur (*Tipster*),
 - Le *track*.
- Réécrire une base en mongoDB avec la structure *horse* comme clef puis *Track* comme clef et enfin la liste des pronostics.

Livrables attendus

Écrire une fonction python qui donne pour deux parieurs tous les chevaux qui ont été parié par l'un comme par l'autre.

Écrire une fonction python qui pour un cheval calcule le nombre de paris positif et le ratio paris positifs sur nombre de paris.

Rajouter des liens sur la base neo4j entre les pronostiqueurs et les champs de courses avec le nombre de pronostics en attribut.

N'hésitez pas à nous montrer toutes autres informations ou fonctions utiles sur ces données.

Documents attendu :

- Le code complet transférer sur le site github, celui-ci doit marcher en démarrant un docker-compose et en lançant un minimum de commande, voir aucune commande.
- Un Readme clair spécifiant :
 - La structure des bases de données.
 - L'utilité de chaque script et leur fonctionnement.

- Les commandes à lancer pour exécuter les différentes fonctions pythons.
- Une présentation détaillant :
 - le set de données.
 - les bases de données et leur structure.
 - les réponses aux livrables.
 - d'autres choses intéressantes s'il y en a.