Jeu d'échec

ANALYSE DE DONNÉES

PAR : ELYES KACEM

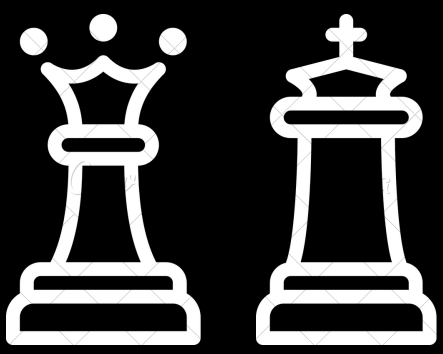
1 ING 1

CONTENTS

SOMMAIRE

DESCRIPTION DU PROJET

TRAVAIL RÉALISÉ



DESCRIPTION DU PROJET

01

Titre : Analyse sur des parties de jeux d'échec d'un fichier contenant **1 MILLION DE PARTIES! (1,000,056 parties)**

02

Lien du projet :

https://colab.research.google.com/drive/1sPT1RBTIeMzuhzL7ig_-O_LGHPCdghdV?usp=sharing

Lien data set :

https://database.lichess.org/standard/lichess_db_standard_rated_2014-09.pgn.bz2

03

Description des colonnes de data set :

- **Event :** le type du partie
- **Site :** identifiant du partie
- **Black_id:** identifiant du joueur noir
- **Black_id :** identifiant du joueur blanc
- **Result:** le résultat de la partie : 0-1 : le noir est le gagnant, 1-0 : le blanc est le gagnant, 1/2-1/2 : Null.
- **UTCDate :** la date de la partie
- **UTCTime :** le temps du départ de la partie.
- **White_ration :** Score du blanc.
- **Black_ration** : Score du noir.
- **WhiteRaitingDiff:** les points gagnant du blanc.
- **BlackRaitingDiff:** les points gagnant du noir.
- **Opening :** l'ouverture de la partie.
- **ECO** : classification des ouvertures.
- **TimeControl :** temps maximum de la partie
- **Termination :** Type de terminaison de la partie : normal, time forfeit(temps épuisé), rule infraction, Abandonned.
- **Translation** : les mouvements dans la partie.

04

Objectifs et Questions de l'analyse :

-Préparer les données (elle était dans une mauvaise format), faire un model et choisir un model plus précis (J'ai utilisé la méthode de Tree classification) et enregistrer le nouuveau fichier et le model dans mon drive.

j'ai fait des études sur la probabilité de gagner la partie et d'avoir une partie null pour chaque joueur, nombres de parties pour chaque type de terminaison, des parties jouéee par la journée et par les heures, les meuilleur ouverture pour chaque joueur, faire des graphes sur la densité de scores en général pour les noirs et les blancs et sur le temps de controle pour les parties.

TRAVAIL RÉALISÉ

01

Description des méthodes d'analyse et d'apprentissage adoptées :

Pour faire le model d'apprentissage, j'ai utilisé la méthode de Tree classification pour préviser la fin de la partie (si c'est le noir qui a gagné : 0, ou le blanc : 1, ou la partie est null: 2) et j'ai subdiviser les données en 2 groupes : données d'apprentissage (80%) et données de teste (20%). Cette méthode permet de faire une arbre de choix, en basant sur les données d'apprentissage. Mais ce type d'apprentissage fonctionne seulement sur les données réelles ou entier. Or j'ai des données de type chaine de caractère, donc j'ai essayé de coder ces données par un clé de type réelle. Après la construction de l'arbre, j'ai tester la précision de ce model en prévisionant les données de test et comparer les résultat obtenu par les valeurs réelles.

Pour l'analyse, j'ai essayer d'organiser les données dans un fichier CSV (car les données était sous une forme indéterminée), puis, j'ai éléminer les lignes conenants les valeurs NaN, et j'ai sauvegarder le nouveau fichier dans le drive. Durant l'analyse, j'ai fais des graphes sur des différents informations en utilisants des différents bibliothèque et modules que vous le trouverez dans les parties suivants.

02

Les bibliothèque python importés :

- **drive** : pour l'intéraction avec mon drive.
- **csv** : pour construire un fichier csv (contenant les données formater)
- **DecisionTreeClassifier** : pour construire le model d'apprentissage
- **train_test_split** : pour séparer les données d'apprentissage et les données du teste
- **accuracy_score** : Pour calculer le score d'accurance.

03

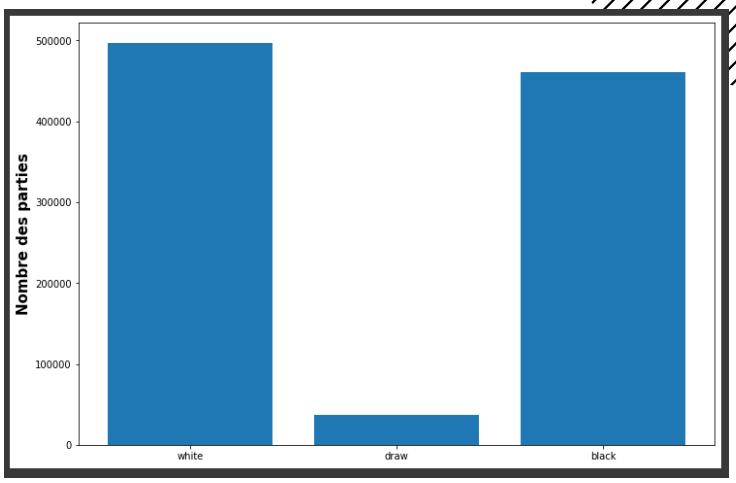
Les modules implémentés :

- **seaborn, plot**
- **matplotlib.pyplot**
- **pandas**
- **sklearn.tree**
- **sklearn.model_selection**
- **sklearn.metrics**
- **joblib**
- **google.colab**

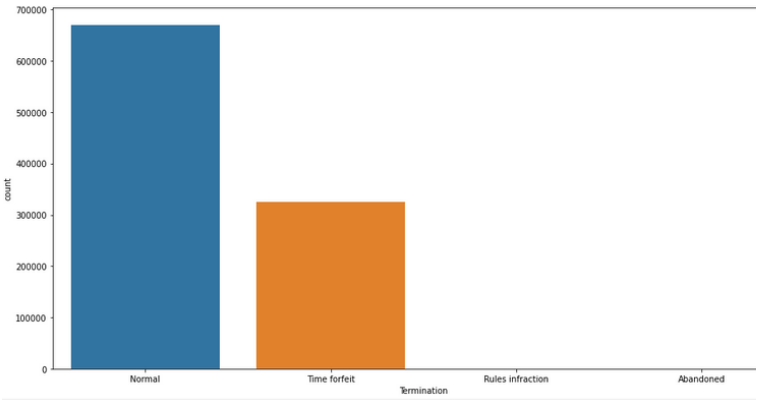
04

Les résultats obtenus :

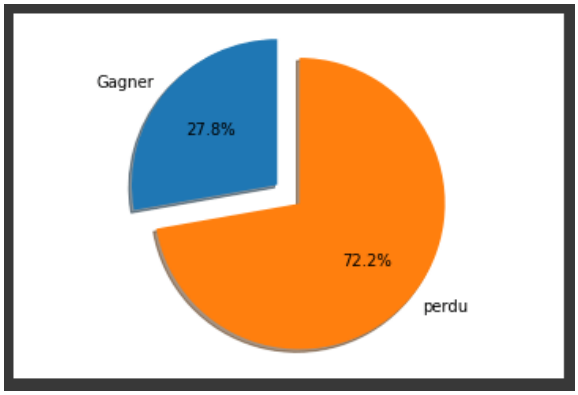
La fin des parties de notre échantillon :



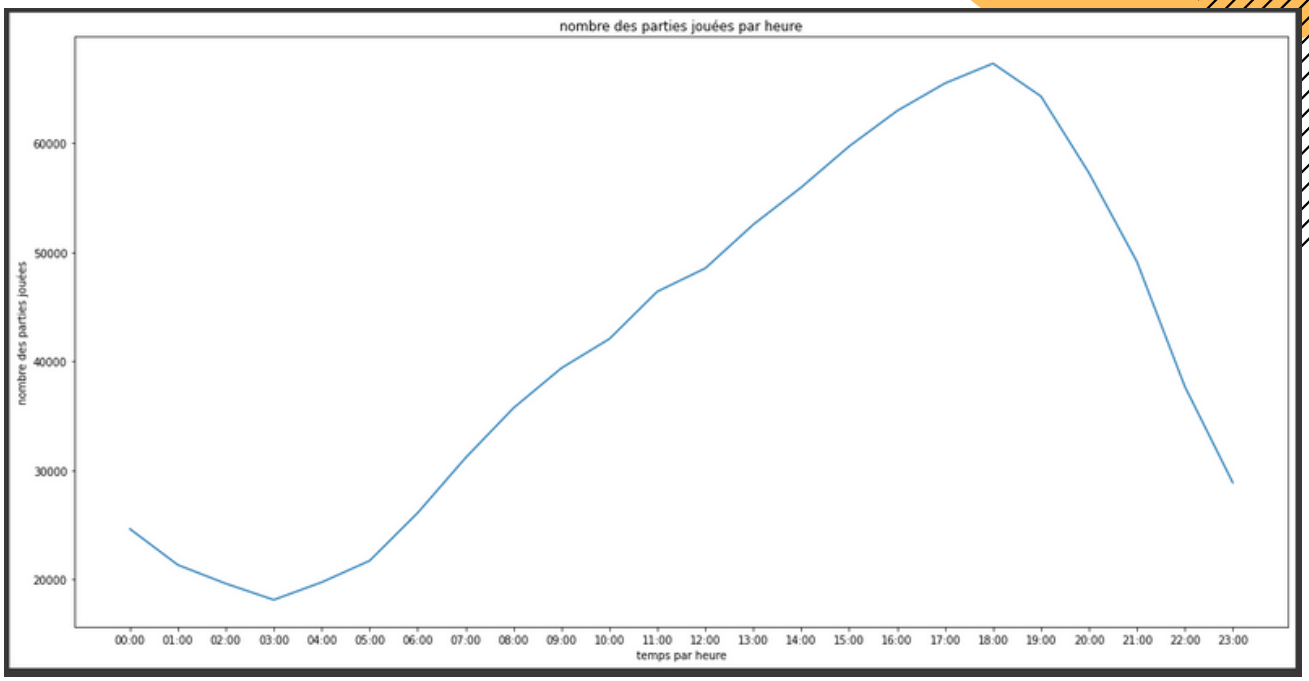
la manière de terminaison pour les parties jouées :



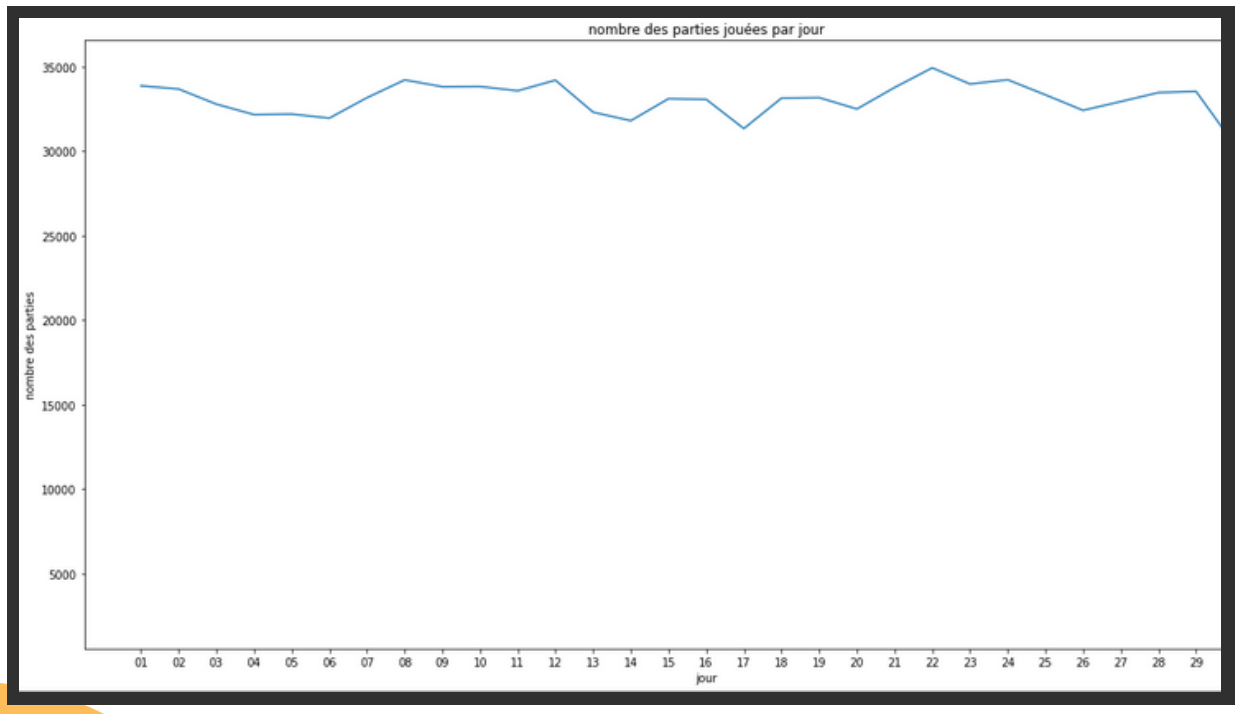
La probabilité de gagner un joueur ayant un score supérieur à votre score :



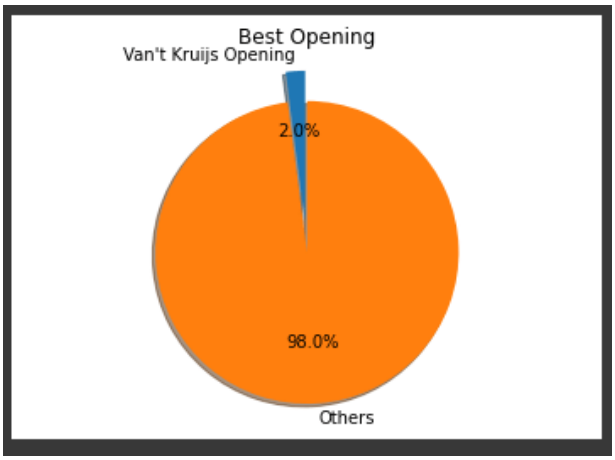
Nombre des parties jouées par heure :



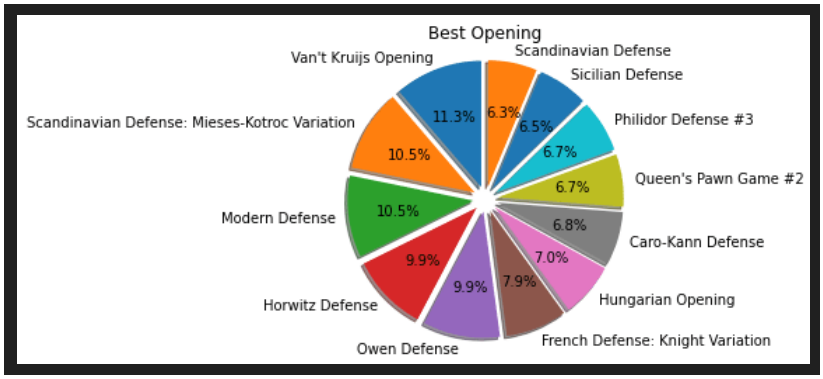
Nombre des parties jouées par jour :



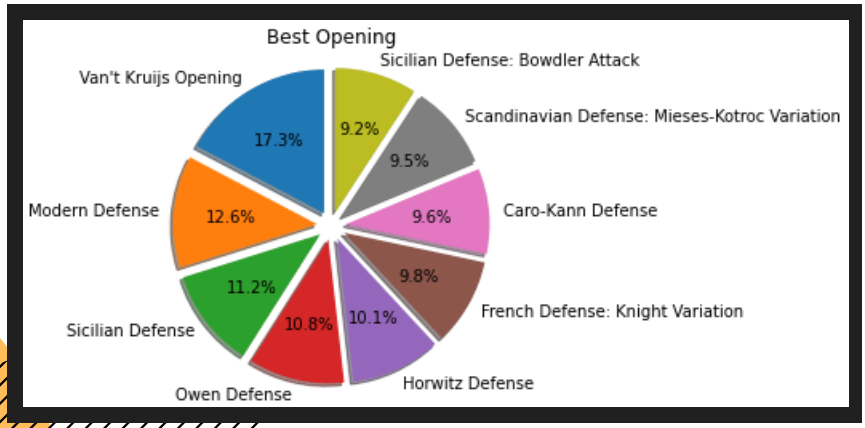
Si vous jouez par les blanc, la meuilleur ouverture de jeu est :
" Van't Krujis Opening "



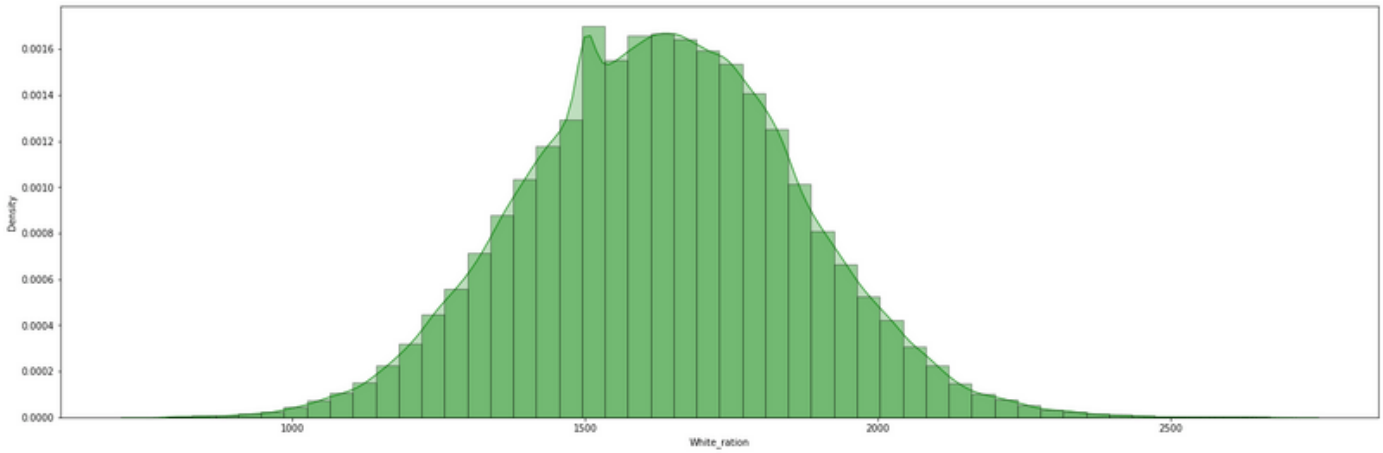
Les 12 meuilleur ouverture pour les blancs :



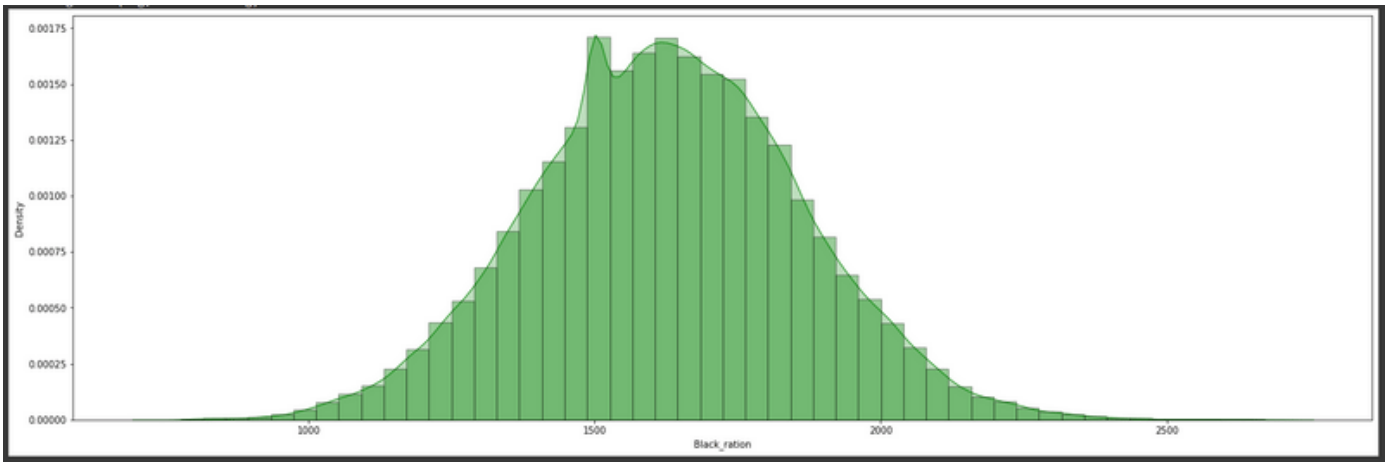
Les meuilleurs ouvertures pour les noirs :



La densité de scores des joueurs blanc
pour cet échantillon :



La densité de scores des joueurs noirs
pour cet échantillon :



J'ai essayé de résumer mon projet dans ce fiche du projet. Pour savoir plus de détails sur le projets, veuillez ouvrir le lien du projet (j'ai travaillé sur Google Collab).
J'espère que vous avez aimez mon projet.

Cordialement,