

# RAPPORT PYTHON - Pierre HOFMANN

## I - Exploration et compréhension

### 1. Importation et inspection

Pour optimiser les performances lors de l'import, j'ai utilisé les type int et float 32 a la place de 64 pour réduire la mémoire utilisé. J'ai également utilisé gc.collect pour vider la mémoire.

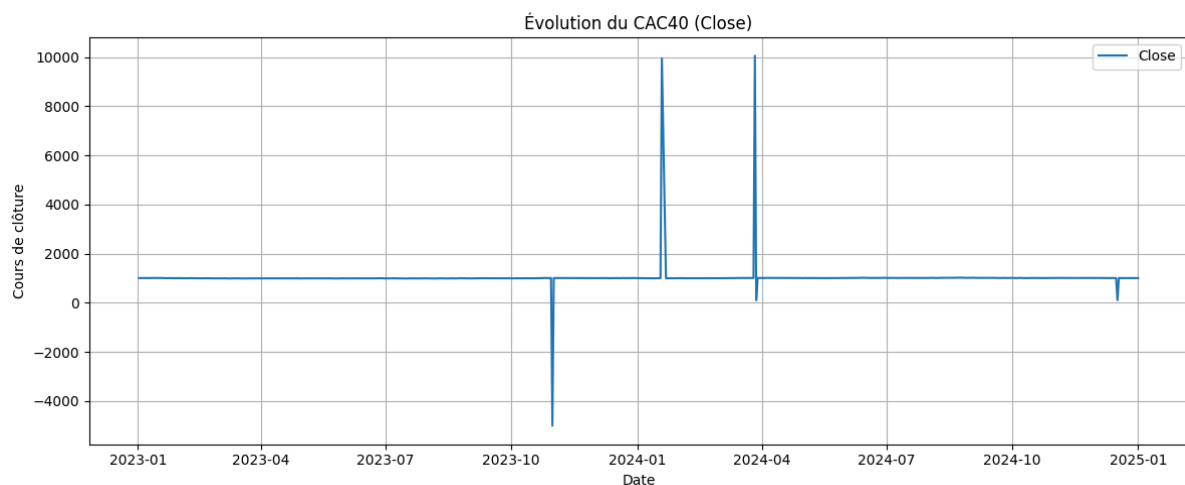
Les colonnes sont globalement propres et bien nommées. Il y a cependant quelque valeur manquant sur certaines lignes.

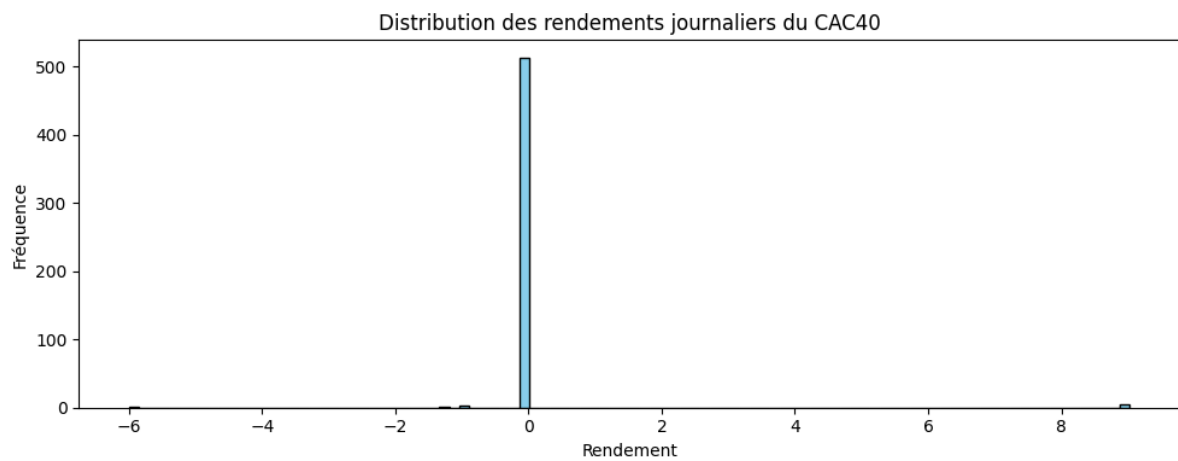
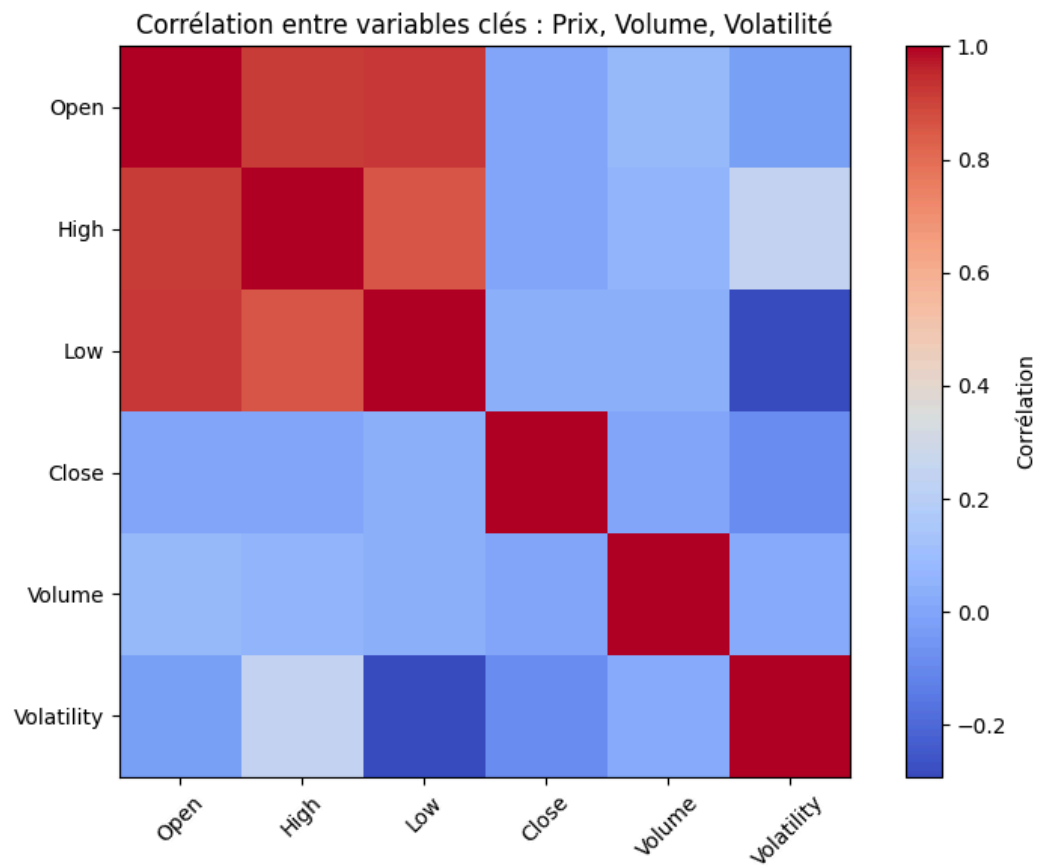
### 2. Nettoyage et mise en forme

J'ai choisis de remplir les valeurs manquantes des colonnes Open, High, Low, Close avec la dernière valeur connue pour éviter d'avoir à supprimer une ligne entière et de garder la continuité des dates. Ça permet donc de ne pas supprimer des informations d'autres colonnes qui peuvent être utiles. Le volume manquant a été remplacé par 0.

La conversion des colonnes numériques en float32 et int32 a réduit la mémoire utilisée par les colonnes financières d'environ 50%. La suppression des doublons a légèrement réduit le nombre de lignes et donc la mémoire totale.

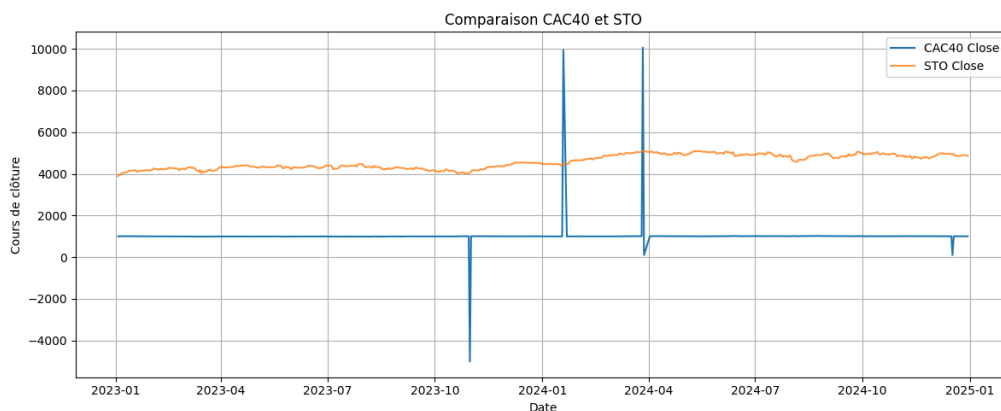
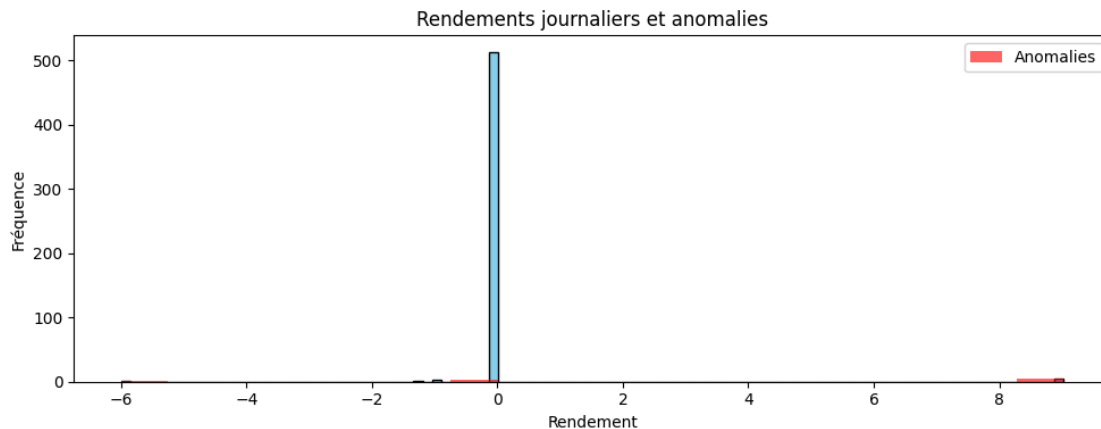
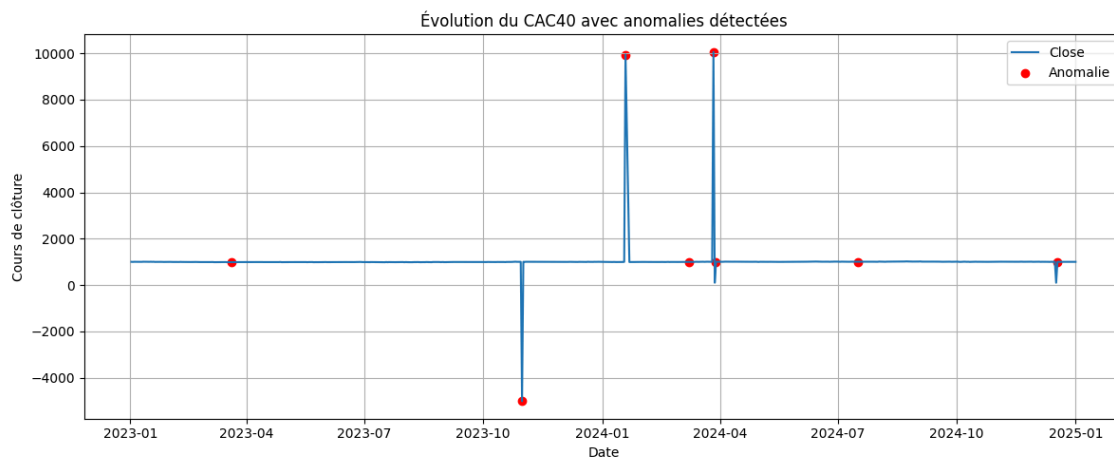
### 3. Analyse visuelle de base via graphiques (Matplotlib)





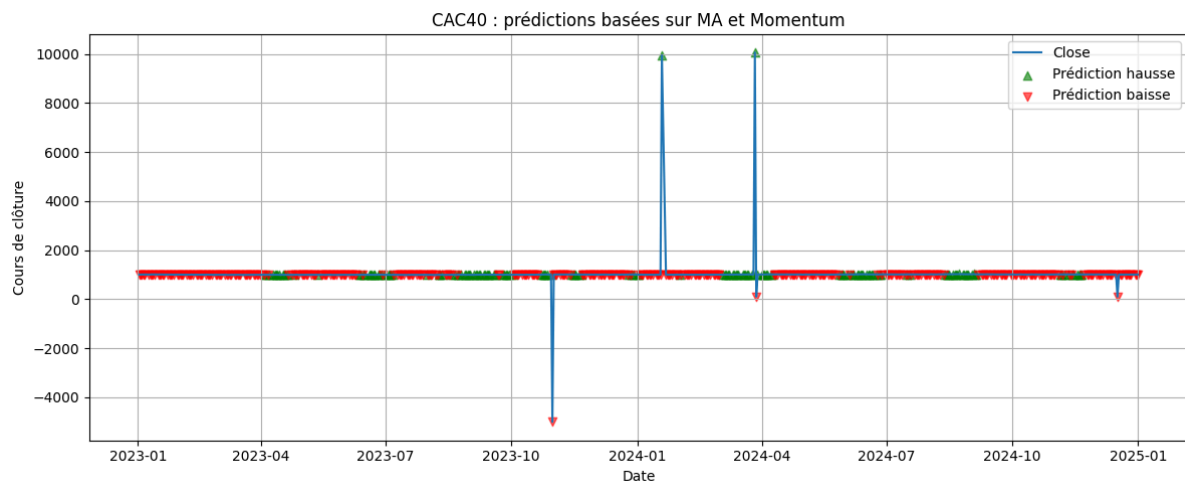
## II - Détection d'anomalies

La méthode choisie pour la détection des anomalies repose sur l'utilisation du z-score appliqué aux rendements journaliers (Return) et à la volatilité (Volatility) du CAC40. Tout point dont la valeur absolue du z-score dépasse 3 est considéré comme une anomalie. Cette approche statistique, simple mais efficace, permet d'identifier rapidement les valeurs extrêmes dans les données. Elle est utile pour repérer les journées où les variations de cours ou de volatilité sont très inhabituelles.

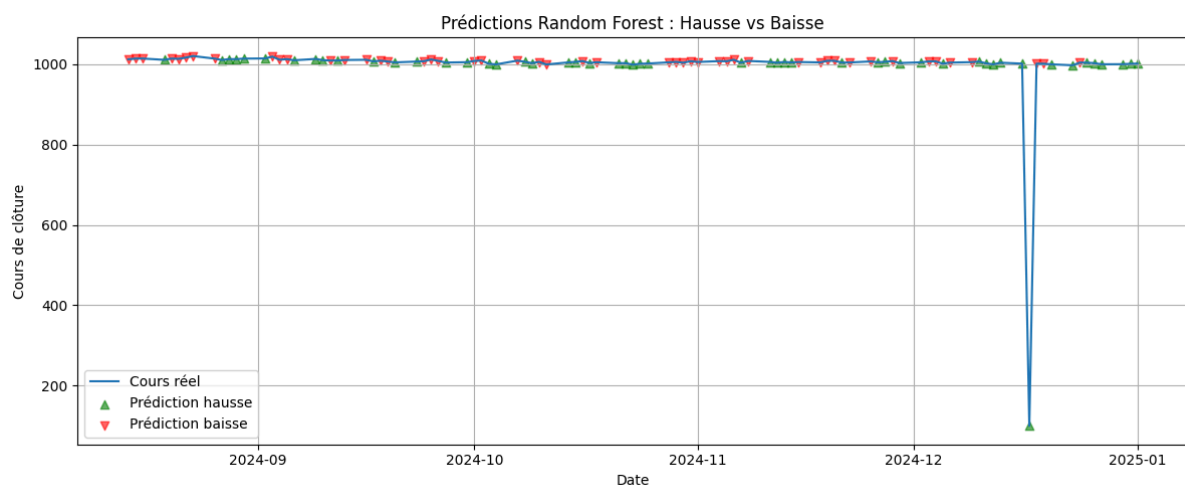


### III - Phase d'analyse statistique

Le modèle simple basé sur les moyennes mobiles et le momentum a une précision proche de 50 %, ce qui montre qu'il ne prédit pas mieux qu'un tirage au sort. Cela montre la complexité des mouvements du CAC40 et la nécessité d'avoir d'autres sources de données pour pouvoir faire des prévisions plus fiables.



### IV - Prédiction avec du Machine Learning



Toutes les données / graphiques présent sur le rapport ont été obtenues grâce au script python et peuvent donc être régénérées avec l'exécution du script.