

Mini-projet Bitcoin — Rapport

Cours 8PRO408 — Outils de programmation pour la science des données. Jeu de données Kaggle `btcsd_1-min_data.csv`, environ 7,3 millions de lignes entre janvier 2012 et décembre 2025.

Préparation et qualité

J'ai commencé par charger le fichier minute en forçant le passage du timestamp Unix en index temporel. Les contrôles de base (`info`, `describe`) montrent cinq colonnes float32, donc l'empreinte reste correcte malgré la taille. Le data set monte à 7 332 157 lignes sans valeur manquante ni doublon. Il reste un seul saut de plus d'une minute, ce qui ressemble à une pause d'échange plus qu'à un trou grave. Le tri chronologique est indispensable avant de lancer les agrégations.

Séries dérivées

Ensuite j'ai produit trois vues : horaire, journalière et mensuelle, avec les mêmes agrégats (première ouverture, derniers prix, moyenne et écart-type de clôture, cumul de volume). La taille globale devient plus digeste : environ 122 000 points pour l'horaire, 5 093 pour la journée et 168 mois. Je pense que ces tables sont directement exploitables pour un reporting simple ou pour nourrir l'app. À côté de ça, j'ai généré des rendements simples et log, plus deux volatilités glissantes (60 minutes et 1 jour) afin de suivre le stress de marché dans le temps. La corrélation Close/Volume lissée sur quinze minutes reste négative (-0,175), ce qui montre que le couple prix-volume n'est pas trivial.

Lecture des courbes

La grande courbe de clôture confirme une tendance haussière sur dix ans, mais avec des phases très brusques. On voit bien les bull runs 2013, 2017 et 2021, et la période 2022 retombe vivement. Les séries de volatilité indiquent que les mouvements minute étaient énormes au début, puis la dynamique se calme un peu après 2019, même si on observe encore des pics à chaque épisode de crise. Les volumes horaires font ressortir des pointes alignées sur les grands retournements, ce qui colle avec l'idée que les traders se mobilisent quand ça secoue. Je me suis appuyé sur ces observations pour commenter les figures du notebook.

Corrélations et interprétations

Pour aller un cran plus loin, j'ai construit un tableau dérivé quotidien avec `pct_return`, `hl_range`, `co_delta` et un volume standardisé `volume_z`. La heatmap sauvegardée dans `figures/corr_heatmap_derived.png` montre des dépendances faibles : les rendements quotidiens collent un peu à `co_delta` (0,51) mais les liens avec le volume restent petits et parfois négatifs. Il me semble que

cela confirme que le volume n'explique pas les mouvements au jour le jour, même si les grosses amplitudes s'accompagnent d'un volume_z supérieur à la moyenne.

Visualisations et application

Le notebook combine Matplotlib et Seaborn pour les graphiques statiques (courbe globale, zoom bull run 2017, histogramme des log-returns, suivis de volatilité, heatmap OHLCV). Plotly s'occupe de la courbe quotidienne interactive. Côté application, j'ai monté un tableau de bord Streamlit qui charge exactement le même CSV. L'utilisateur choisit la période avec deux sélecteurs de dates et bascule entre plusieurs granularités (1 min à 1 jour). La page affiche ensuite la courbe de prix filtrée, trois métriques simples (dernier prix, min/max, volume total), puis une vue combinée prix/volume où l'axe droit suit le volume agrégé. On termine avec un tableau de statistiques descriptives pour la fenêtre affichée. J'ai lancé l'app localement via `streamlit run app_streamlit.py`, ça se charge sans erreur tant que le fichier est en place et la navigation reste fluide.

Limites et suites

Le data set ne comble pas les rares gaps, donc une future étape serait de les annoter proprement ou de décider d'une interpolation. Je n'ai pas encore ajouté d'indicateurs techniques type RSI ou bandes de Bollinger ; ce serait une bonne extension pour préparer un modèle de prévision. Un autre chantier serait de refaire les mêmes analyses sur des fenêtres ciblées comme le crash COVID 2020 ou le bear market 2022 pour comparer les réactions. Globalement, je considère que le livrable respecte les consignes et fournit une base solide mais assez simple, conforme à ce qu'on attend d'un mini-projet.

Merci d'avoir lu jusqu'au bout, et désolé pour le retard, c'était un projet fun à faire, j'ai bien aimé le sujet.

Vianney Jacquemot :)