

Scraping, Cleaning and Analyzing

Introduction :

L'analyse et la visualisation de données sont des étapes essentielles pour comprendre les tendances, les relations et les insights cachés dans les données collectées. Dans ce projet, nous avons exploré les méthodes et les outils nécessaires pour effectuer une analyse approfondie et créer des visualisations claires et informatives à partir des données de GitHub.

Topographie des données :

Les données GitHub ont été collectées pendant une période d'une année de juin 2022 à juin 2023. Les données comprenaient des informations sur les dépôts, les contributeurs et les contributions... Les données ont été collectées à l'aide de l'API GitHub .

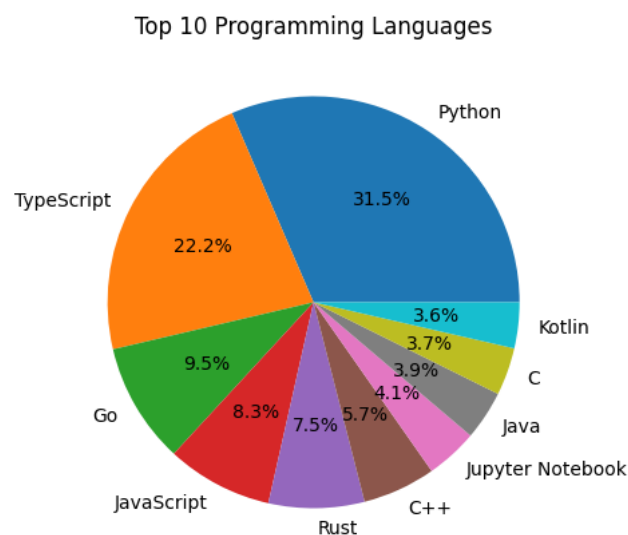
Questions clés :

Pour extraire des insights significatifs des données GitHub, nous nous sommes concentrés sur les questions clés suivantes :

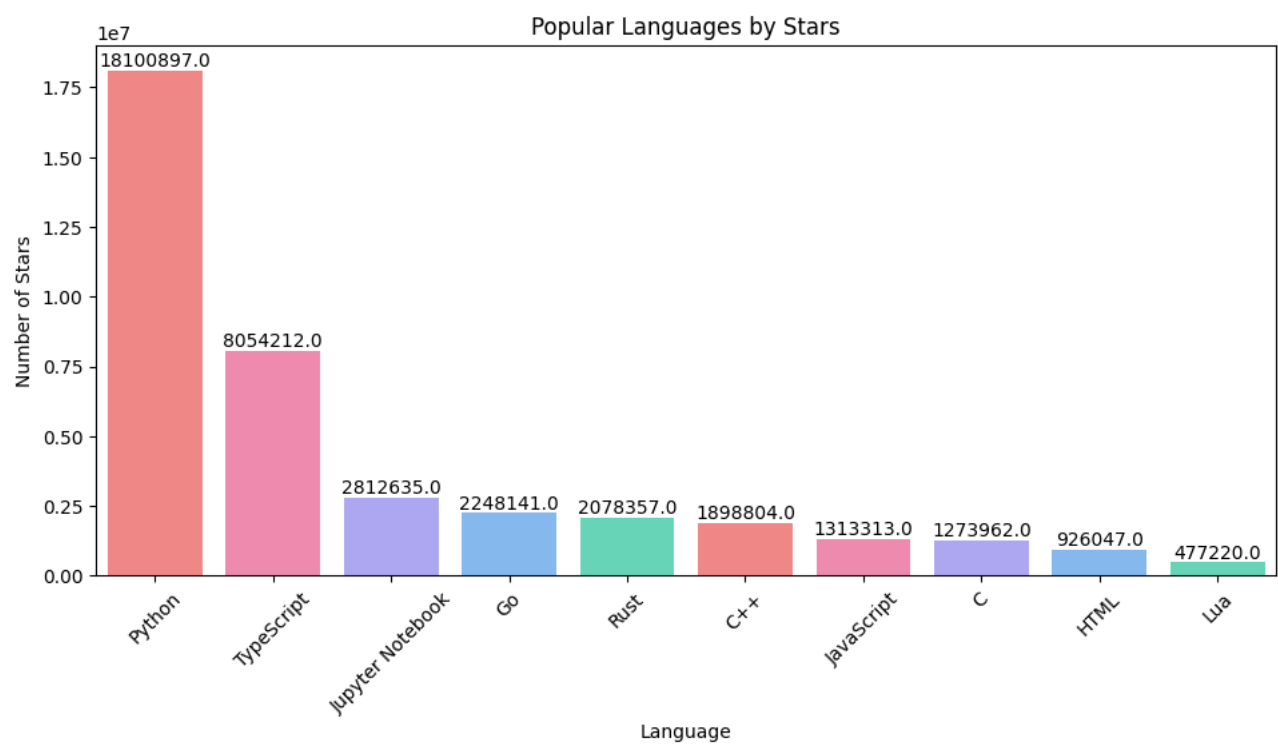
- Quels sont les principaux langages de programmation utilisés dans les dépôts GitHub analysés et quelle est leur répartition parmi les dépôts ?
- Quels sont les 10 langages de programmation les plus populaires avec le plus grand nombre d'étoiles sur les dépôts ?
- Quels sont les 10 langages de programmation les plus populaires avec le plus grand nombre de forks sur les dépôts ?
- Quels sont les 10 langages de programmation les plus populaires avec le plus grand nombre de problèmes ouverts sur les dépôts ?
- Quels sont les 10 langages de programmation les plus populaires avec le plus grand nombre de contributeurs sur les dépôts ?
- Quels sont les 10 contributeurs les plus actifs ?
- Quels sont les 10 dépôts les plus populaires de tous les temps ?
- Quels sont les sujets les plus courants entre tous les dépôts ?
- Comment ont évolué les tendances au fil du temps pour les 10 langages de programmation les plus populaires ?
- Quelle est la corrélation entre les variables 'stars', 'forks', 'open_issues' et 'contributions' ?

Analyse :

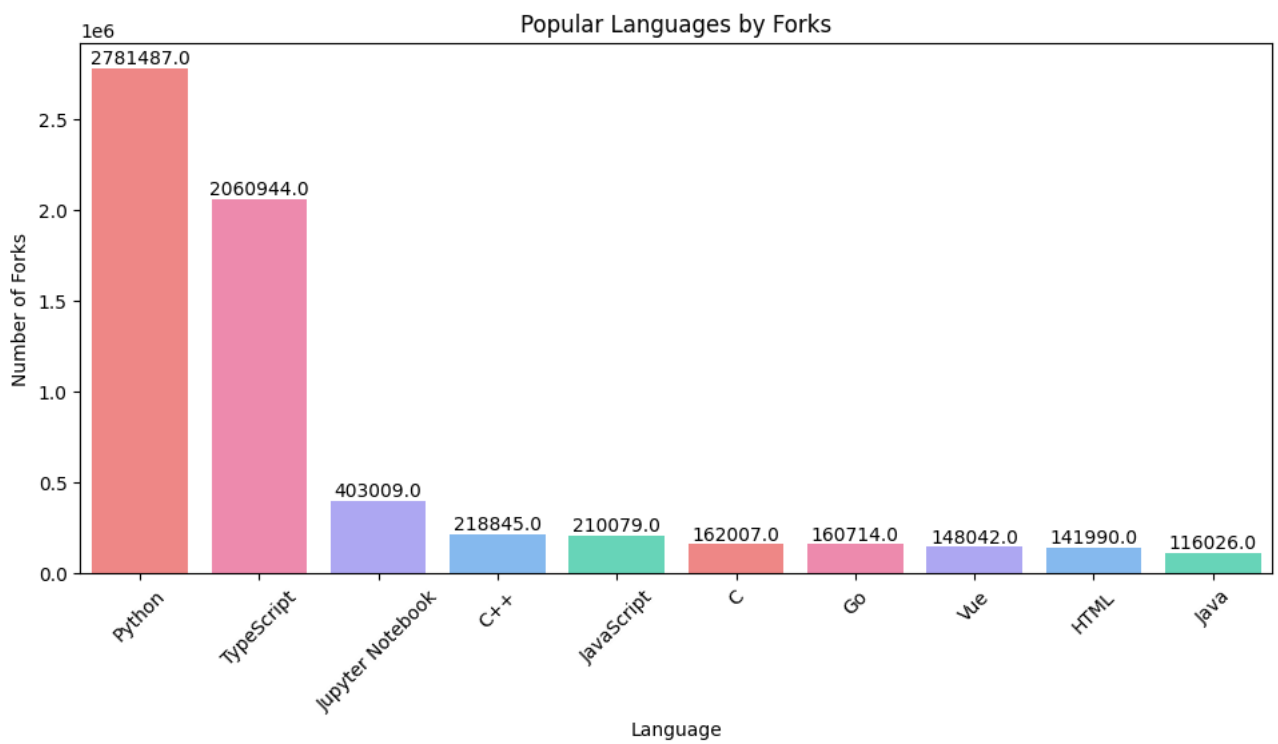
Les principaux langages de programmation utilisés dans les dépôts GitHub analysés et leur répartition parmi les dépôts



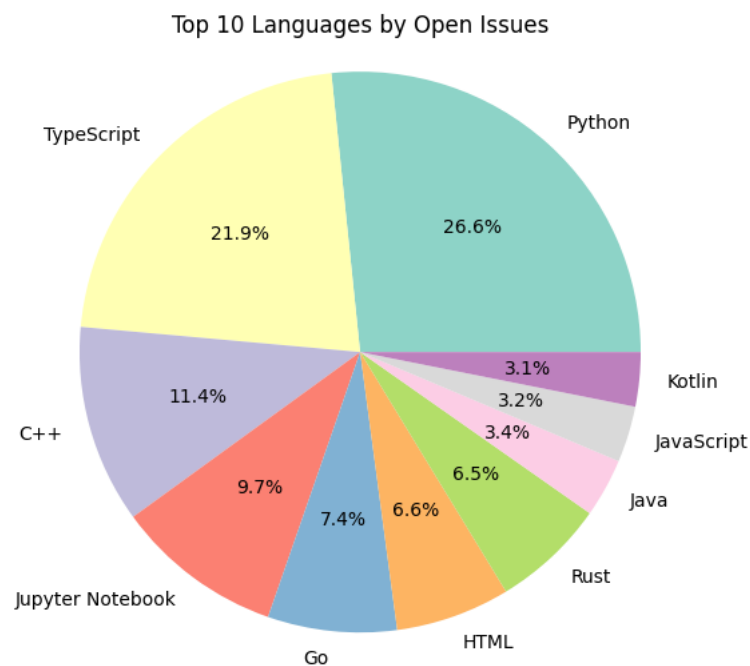
Les 10 langages de programmation les plus populaires avec le plus grand nombre d'étoiles



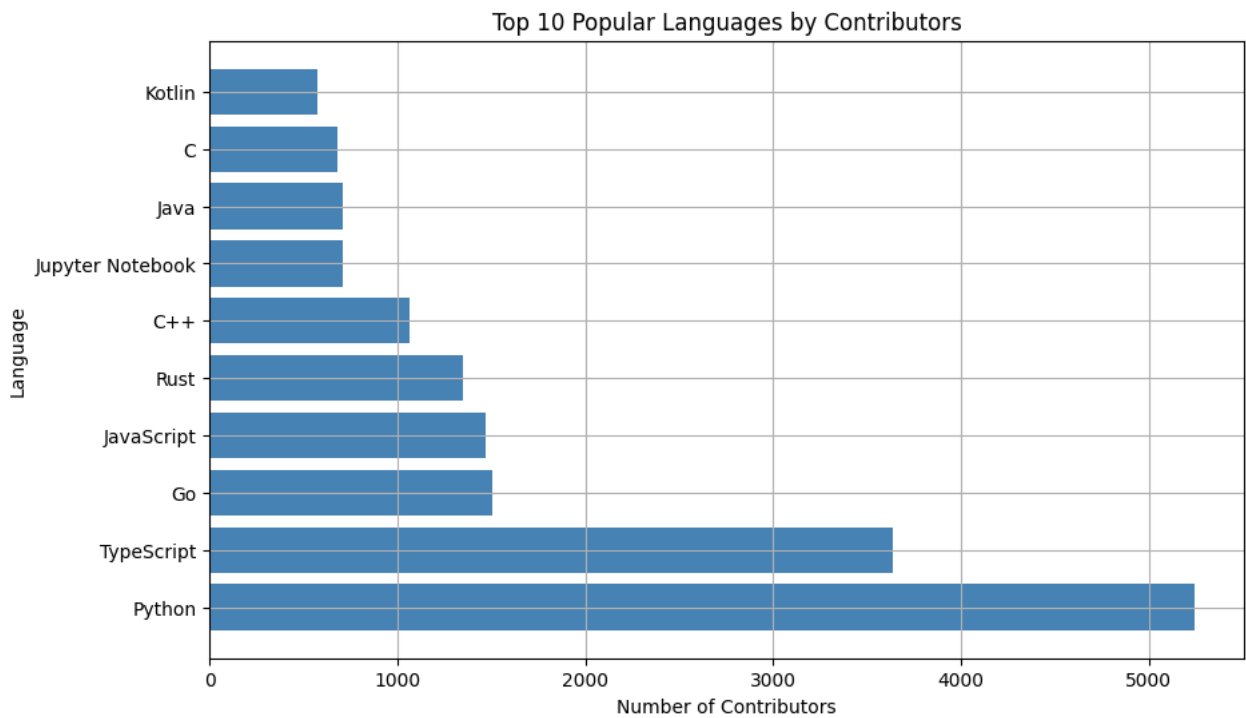
Les 10 langages de programmation les plus populaires avec le plus grand nombre de forks



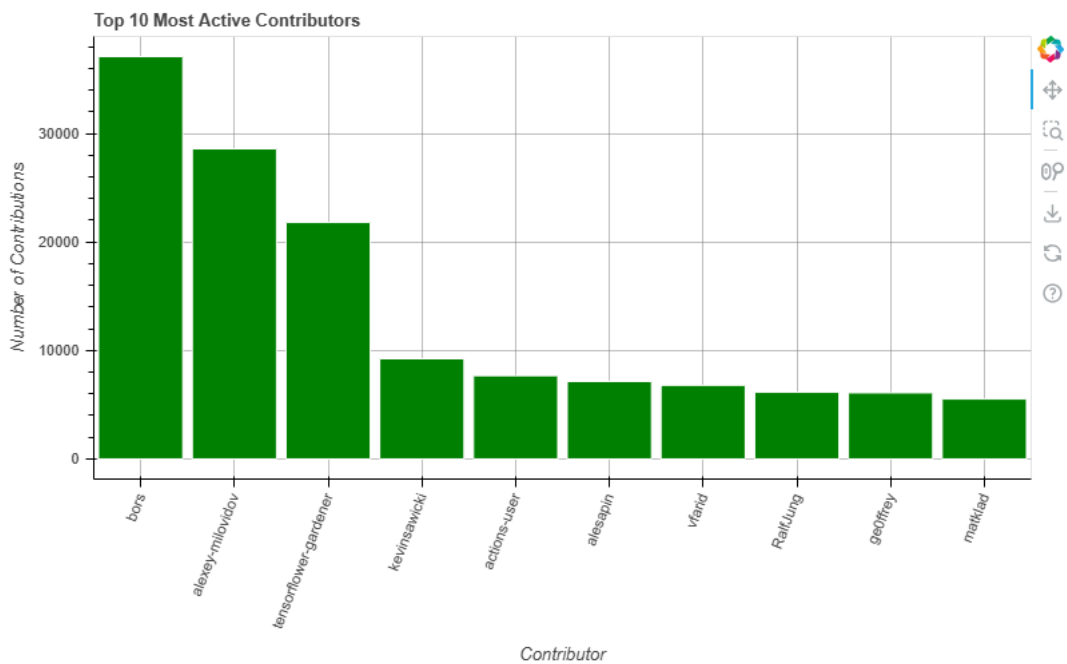
Les 10 langages de programmation les plus populaires avec le plus grand nombre de problèmes ouverts



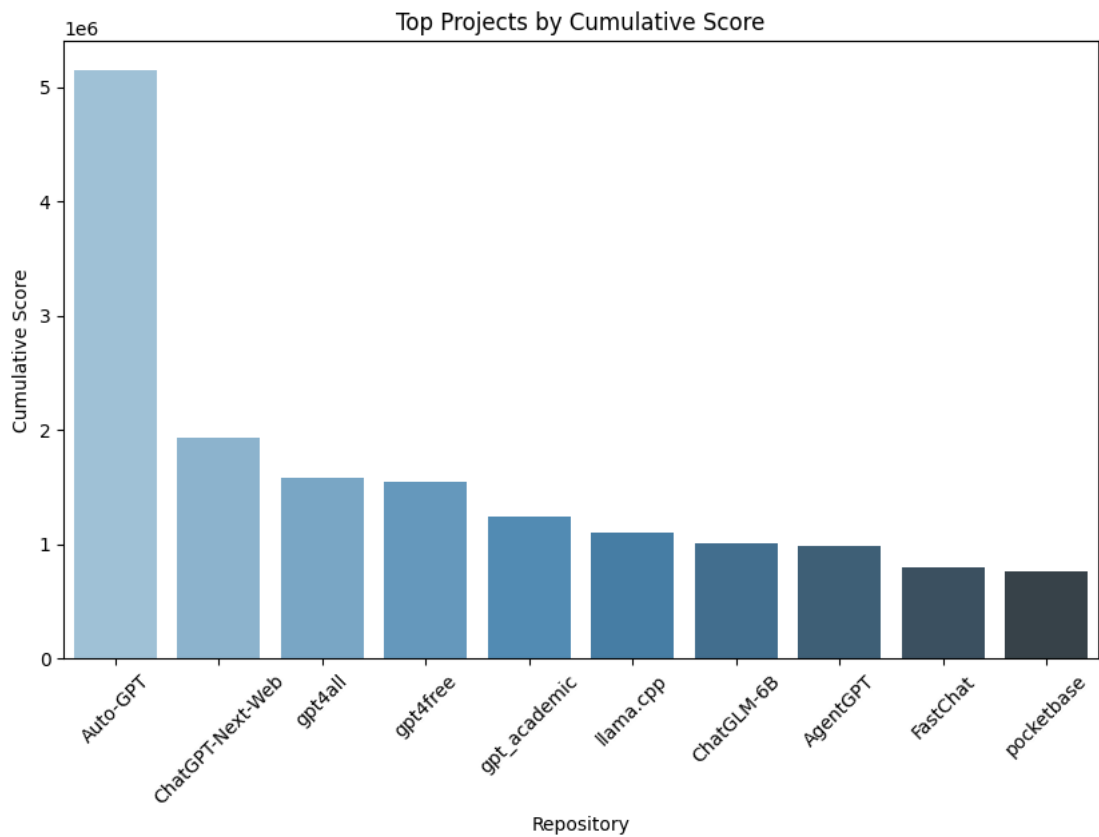
Les 10 langages de programmation les plus populaires avec le plus grand nombre de contributeurs



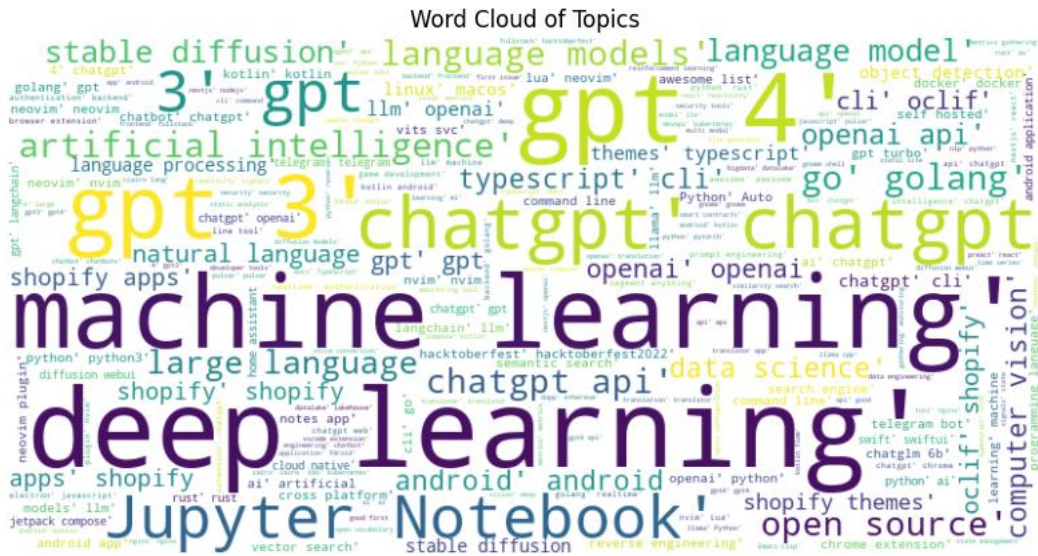
Les 10 contributeurs les plus actifs



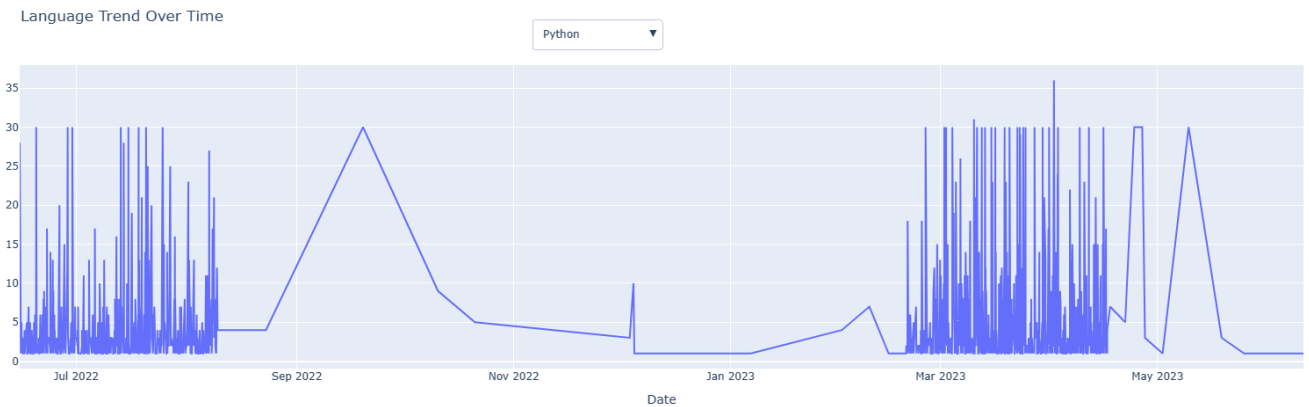
Les 10 dépôts les plus populaires de tous les temps



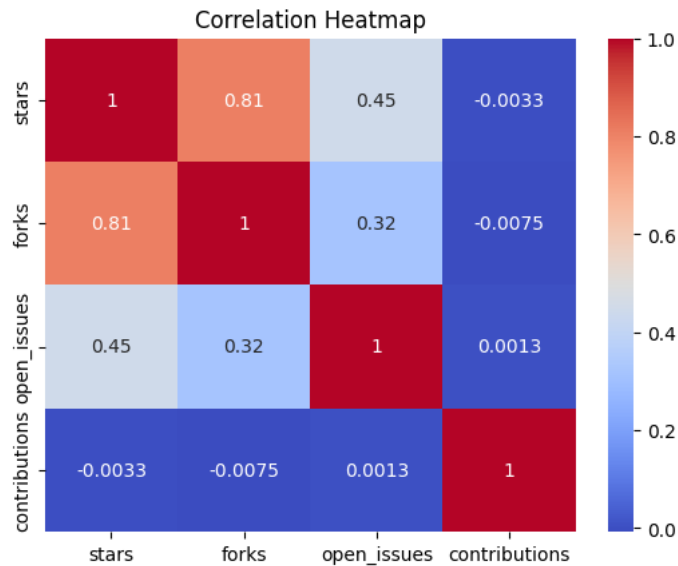
Les sujets les plus courants entre tous les dépôts



Les tendances au fil du temps pour les 10 langages de programmation les plus populaires



La corrélation entre les variables 'stars', 'forks', 'open_issues' et 'contributions'



Conclusion :

En conclusion, l'analyse et la visualisation des données GitHub ont fourni des insights significatifs sur les tendances, les relations et la popularité des dépôts. Les insights peuvent être utilisés pour améliorer le processus de développement et identifier des domaines d'amélioration. Les visualisations créées pour l'analyse étaient claires et informatives, communiquant efficacement les résultats de manière compréhensible.