

# Collecte automatisée de données web

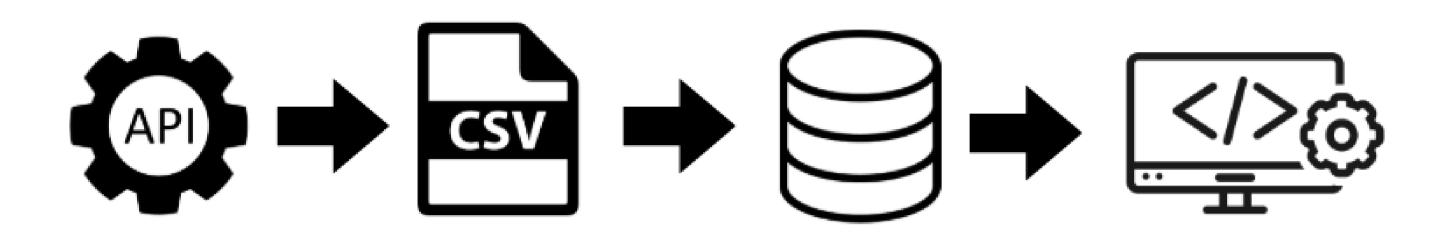
Romain Lesueur Sidy Diop Courteney Saint-Hubert

BUT Science des Données - Parcours Visualisation et Conception d'Outils Décisionnels



#### Introduction

Dans le domaine des marchés financiers, l'analyse des données boursières et des actualités influençant les entreprises est cruciale. « BourseTrack » est une plateforme qui centralise et structure ces informations afin de suivre les performances financières des grandes entreprises sur les 100 derniers jours ouvrés. Grâce à une interface simple et intuitive, le projet permet aux utilisateurs d'examiner les fluctuations des données et des actualités associées pour une analyse rapide des tendances économiques.



En résumé, l'objectif est de développer une page web simple et intuitive permettant de suivre les performances financières des grandes entreprises sur les 100 derniers jours ouvrés.

#### Répartition des tâches

Afin d'assurer une gestion optimale du projet, les responsabilités ont été attribuées en fonction des compétences et des domaines maitrisés de chaque membre.

- Romain : Gestion des APIs et collecte des données boursières (récupération des prix, volumes, actualités).
- Sidy: Gestion de la base de données (stockage, nettoyage, calcul des indicateurs financiers).
- Courteney : Développement du site web (design, interface utilisateur, affichage des données et actualités).

La répartition des tâches est essentielle pour garantir une organisation fluide et une gestion efficace du projet.

#### Définition

Ci-dessous, vous trouverez des définitions des termes techniques clés utilisés dans ce projet, vous permettant ainsi de mieux comprendre les concepts abordés dans ce poster.

- API (Application Programming Interface) : Une interface qui permet à différentes applications de communiquer entre elles. Dans ce projet, elle est utilisée pour récupérer des données boursières et des actualités financières.
- Base de données : Un système de stockage qui organise et conserve les données de manière structurée. Elle permet de gérer les informations financières et de faciliter l'accès aux données par l'application.
- Rate-limit : Technique utilisée pour contrôler la fréquence des requêtes envoyées à une API ou à un système, afin d'éviter une surcharge et d'assurer une distribution équitable des ressources.
- Tableau dynamique : Type de tableau interactif qui permet de manipuler, filtrer et analyser les données en temps réel, offrant une grande flexibilité pour visualiser des informations complexes.
- HTML : Langage de balisage utilisé pour structurer et organiser le contenu des pages web, comme les textes, images, liens, et multimédias.

#### **API**

Les APIs sont essentielles pour collecter les données nécessaires au projet, en fournissant des informations fiables sur les marchés financiers et les actualités économiques. Deux APIs principales ont été sélectionnées pour répondre aux besoins du projet.

- 1. Alpha Vantage (https://www.alphavantage.co/) : Permet d'accéder aux données boursières des entreprises, telles que les prix d'ouverture, de clôture, les volumes échangés et d'autres indicateurs sur les 100 derniers jours ouvrés.
- 2. News API (https://newsapi.org/) : Fournit des actualités économiques et financières, mettant en lumière les événements ayant un impact sur les entreprises suivies.
- 3. **Limitations et configuration** : Ces APIs imposent des restrictions sur le nombre de requêtes par période (*rate-limit*). De plus, chaque API nécessite une clé d'accès unique, qui doit être obtenue et correctement configurée avant de pouvoir les utiliser.

#### **CSV**

Le format CSV (*Comma-Separated Values*) est utilisé pour stocker localement les données récupérées via l'API. Cela permet de limiter les requêtes répétées à l'API et d'éviter les restrictions dues au *rate-limit*. De plus, ce format simple et universel est idéal pour conserver et analyser les informations financières.

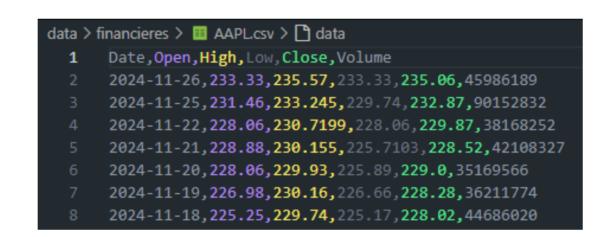


Figure 1. Exemple de fichier CSV.

Les données sauvegardées en CSV permettent :

- Un accès rapide sans solliciter l'API à chaque requête.
- Une compatibilité avec de nombreux outils d'analyse.
- Une gestion simplifiée des informations boursières.

Ce choix garantit **efficacité et flexibilité** dans le traitement des données.

#### Base de données

La base de données utilisée pour ce projet est conçue pour organiser efficacement les informations financières et les relier aux entreprises suivies. Elle repose sur deux tables principales "Entreprises" et "Donnée financière", qui centralisent respectivement les informations générales sur les entreprises et leurs données boursières.

Champ	Description
ID	Identifiant unique
Symbole	Symbole boursier (référence à "Entreprises")
Secteur	Secteur d'activité
Date	Date de la donnée
Open	Prix d'ouverture
High	Prix le plus élevé
Low	Prix le plus bas
Close	Prix de clôture
Volume	Volume échangé

Table 1. Table "Donnée financière"

Les tables sont reliées par la colonne **Symbole**, qui établit une correspondance entre les données boursières et l'entreprise associée. Cette structure simplifie l'accès et l'analyse des informations nécessaires au suivi des performances.

### Création du tableau dynamique

Ce code HTML génère un tableau dynamique affichant les indicateurs financiers pour chaque entreprise. Il comprend des informations telles que les prix moyens d'ouverture et de fermeture, la volatilité, le volume moyen, la tendance globale et le rendement moyen journalier. Le tableau est conçu pour être interactif, permettant aux utilisateurs de trier les colonnes en fonction de leurs besoins en cliquant sur les en-têtes des colonnes.

```
ableau des indicateurs financiers -->
e border="1" id="tableau-indicateurs">
    Entreprise<span class="sort-indicator"></span>
    Prix Moyen Fermeture<span class="sort-indicator"></span>
    Volatilité Moyenne<span class="sort-indicator"></span>
    Volume Moyen<span class="sort-indicator"></span></th</pre>
    Tendance Globale<span class="sort-indicator"></span>
    Rendement Moyen Journalier<span class="sort-indicator"></span>
 {% for indicateur in indicateurs %}
   {{ indicateur.nom }}
   {{ indicateur.prix_moyen_ouverture | round(2) }}
   {{ indicateur.prix_moyen_fermeture | round(2) }}
    {{ indicateur.volatilite_moyenne | round(2) }}
    {{ indicateur.volume moven | round(2) }}
   {{ indicateur.tendance_globale | round(2) }}
   {{ indicateur.rendement moyen journalier | round(2) }}
{% endfor %}
```

Ce tableau permet d'afficher de manière claire et structurée les informations financières essentielles. Il présente des données arrondies à deux décimales pour améliorer la lisibilité. Les utilisateurs peuvent également trier les colonnes en fonction de différents critères.

## Rendu du tableau dynamique sur la page Web

Afin de donner une vision concrète du projet, voici une capture d'écran du rendu du site web que nous avons développé. Cette interface permet de visualiser les indicateurs financiers de manière claire et intuitive, facilitant ainsi l'analyse des données pour les utilisateurs.

#### Performances Boursières des Grandes Entreprises **Prix Moyen** Prix Moyen **Entreprise** Volume Moyen Tendance Globale Ouverture 225.34 Apple Inc. 225.55 0.11 186.76 186.57 39976756.46 -0.19 -0.09 Amazon.com, Inc Alibaba Group 89.25 89.28 0.05 17744397.36 Holding 0.01 Ford Motor Compan 58310526.3 Microsoft 423.06 422.34 -0.16 20470306.13 Corporation -0.06 Toyota Motor Corp 180.63 360169.53 246.1 245.81 -0.28 -0.06 Tesla, Inc. 92638267.38

Cette image illustre l'interface utilisateur où sont affichés les différents indicateurs financiers. Le design met en avant les données essentielles, telles que les prix moyens d'ouverture et de fermeture, la volatilité, ainsi que les tendances du marché, dans un format lisible et interactif.