



Automatiser l'Investissement : Scraper Yahoo! Finance avec Python

De la donnée brute à l'intelligence financière :
une étude de cas sur les 'Most Active Stocks'.

Un guide visuel pour data scientists et investisseurs modernes.

La Mission : Démocratiser la Donnée Financière

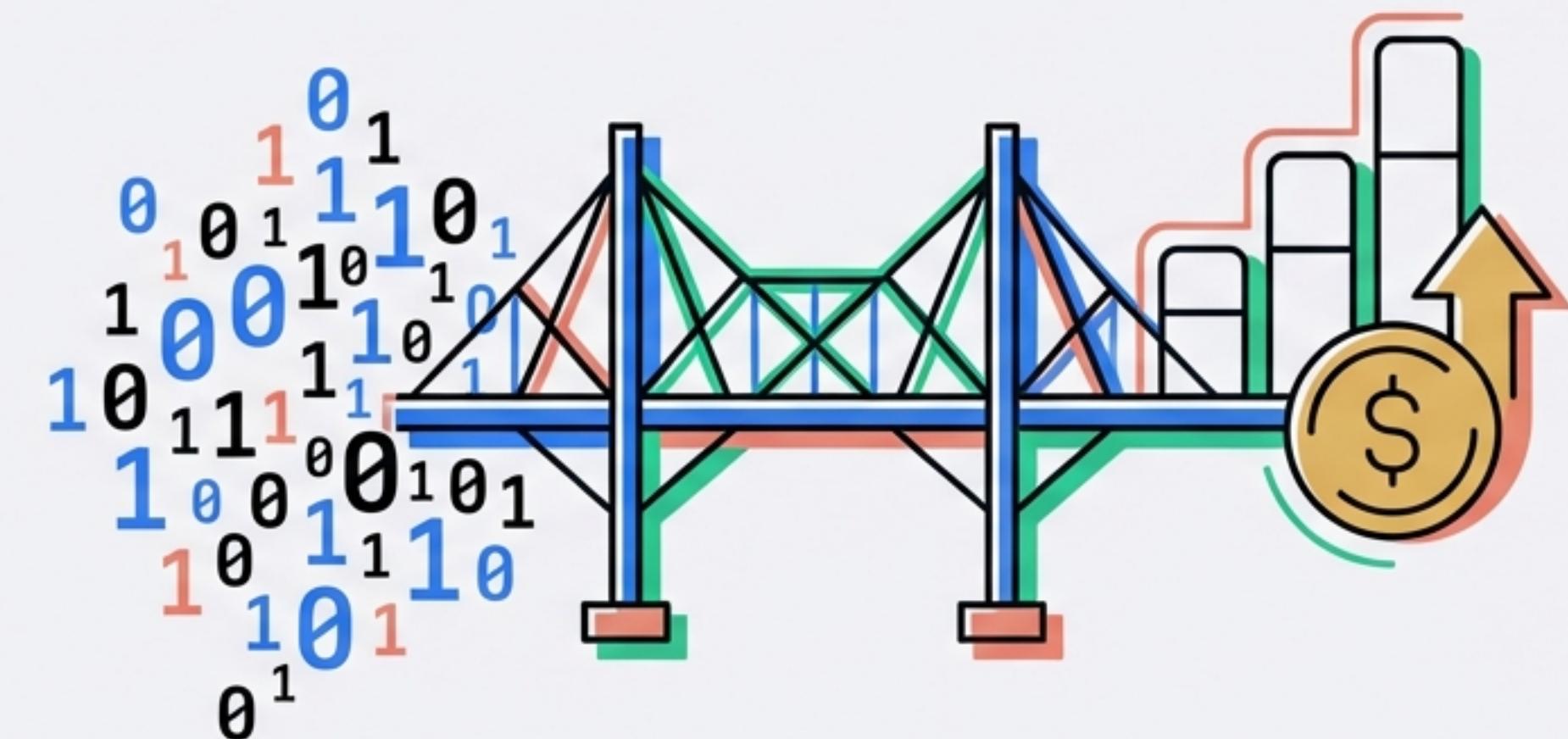
L'investissement est un levier puissant, mais l'information est souvent dispersée.

La Motivation

Favoriser l'autonomie financière et contribuer à la réduction des écarts de richesse par le partage de connaissances technologiques.

La Cible

Les 'Actions les plus actives' (Most Active Stocks) – les 200 titres présentant le volume quotidien le plus élevé sur le marché.



Il est crucial d'apprendre et de partager des connaissances sur la finance pour favoriser l'autonomie financière de chacun.

Le Glossaire de l'Investisseur



Capitalisation Boursière

La valeur totale de l'entreprise (Prix de l'action × Nombre d'actions). Permet de classer les entreprises (Small, Mid, Large caps).



Volume

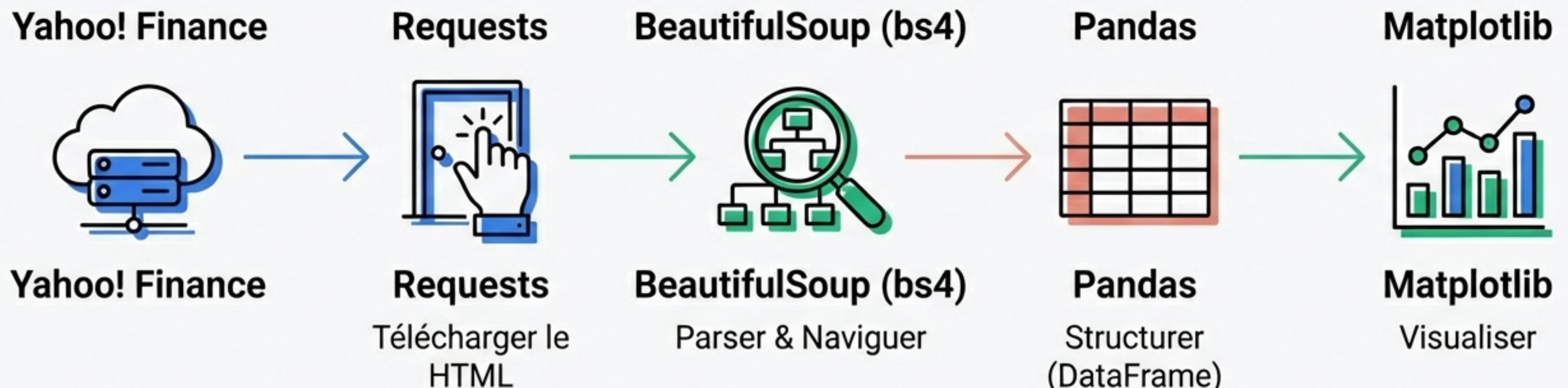
La quantité de titres échangés sur une période donnée. C'est le baromètre de l'activité du marché.



Ratio PE (Price-to-Earnings)

Prix de l'action divisé par les bénéfices. Un ratio faible peut indiquer une "bonne affaire" (sous-évaluation).

L'Architecture Technique



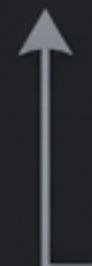
```
import requests  
from bs4 import BeautifulSoup  
import pandas as pd
```

Éthique & Préparation : Le 'Savoir-Vivre' du Web

Avant de récolter, on vérifie les règles du site via `robots.txt`.

```
# HEADERS ÉTHIQUES
# On s'identifie clairement sans se faire passer pour un robot malveillant
HEADERS = {
    'User-Agent': 'Mozilla/5.0... (Student Project; Contact: nenadjojan6@gmail.com)',
    'Accept-Language': 'en-US,en;q=0.9',
}

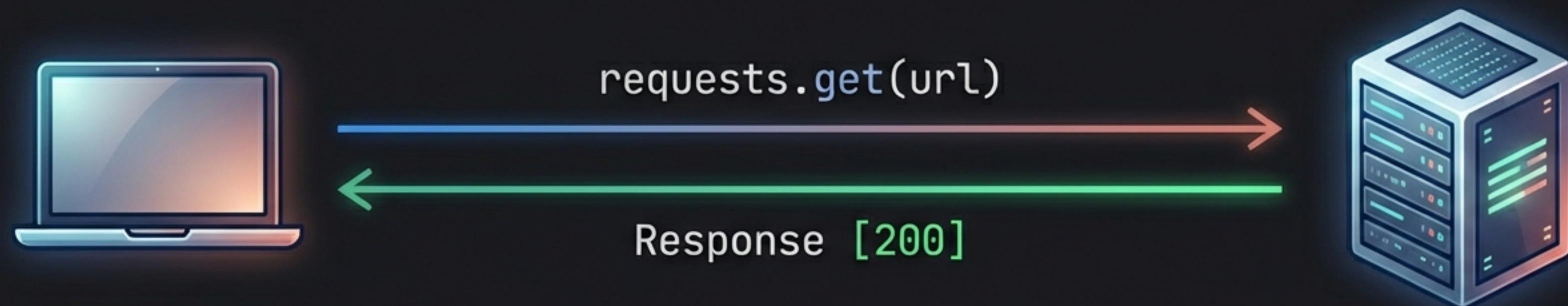
def check_robots_txt(url, user_agent):
    # Vérifie si le scraping est autorisé par le site
    # ... code de vérification ...
```



S'identifier clairement évite d'être bloqué.

✓ Autorisation Robots.txt accordée pour : finance.yahoo.com/most-active

L'Extraction : La Première Connexion



```
url = 'https://finance.yahoo.com/markets/stocks/most-active/' \
response = requests.get(url, headers=headers)
# Status Code: 200 (Succès)
```

- ✓ Si le code est entre 200 et 300, le téléchargement est réussi.

Décrypter la Structure : Parsing HTML

```
Elements Console Sources Network Performance > | ⚙️ ⚙️ X
▼<div class="row yf-1w0dr5b">
  ▶<div class="row yf-1w0dr5b" ta-data">...</div>
  ▶<div class="row yf-1w0dr5b">...</div>
...
  ▼<table class="row yf-1w0dr5b" antwralter="row yw0dr5b" row yf-1w0dr5b1! caibsut°1ve;"> ...
    ▶<thead>
      ▶<tbody>
        ▶<tr class="row yf-1w0dr5b">...</tr>
        ▶<tbody>
          ▶<tr>
...
    ▶<tr class="row yf-1w0dr5b" = $0
      ▶<td class="row yf-1w0dr5b">..</td>
      ▶<td class="row yf-1w0dr5b">..</td>
      ▶<td class="row yf-1w0dr5b">..</td>
      ▶<td class="row yf-1w0dr5b">..</td>
      ....
    </tr>
  </tbody>
</thead>
</div>
</div>
```

```
doc = BeautifulSoup(html)
```

Find <table>

(Le conteneur principal)

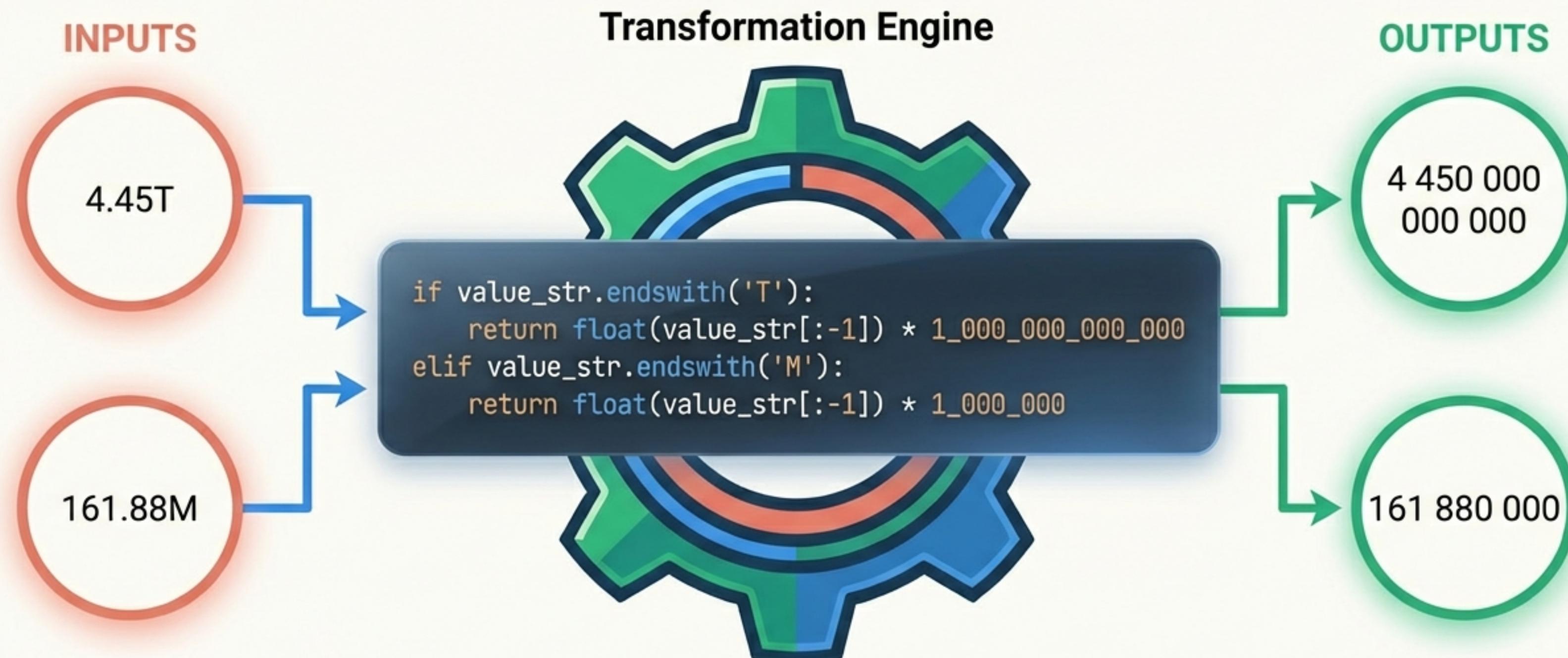
Find_all <tr>

(Les 25 lignes du tableau)

```
tbody_tag = doc.find('tbody')
tr_tags = tbody_tag.find_all('tr')
print(f"Found {len(tr_tags)} rows.")
```

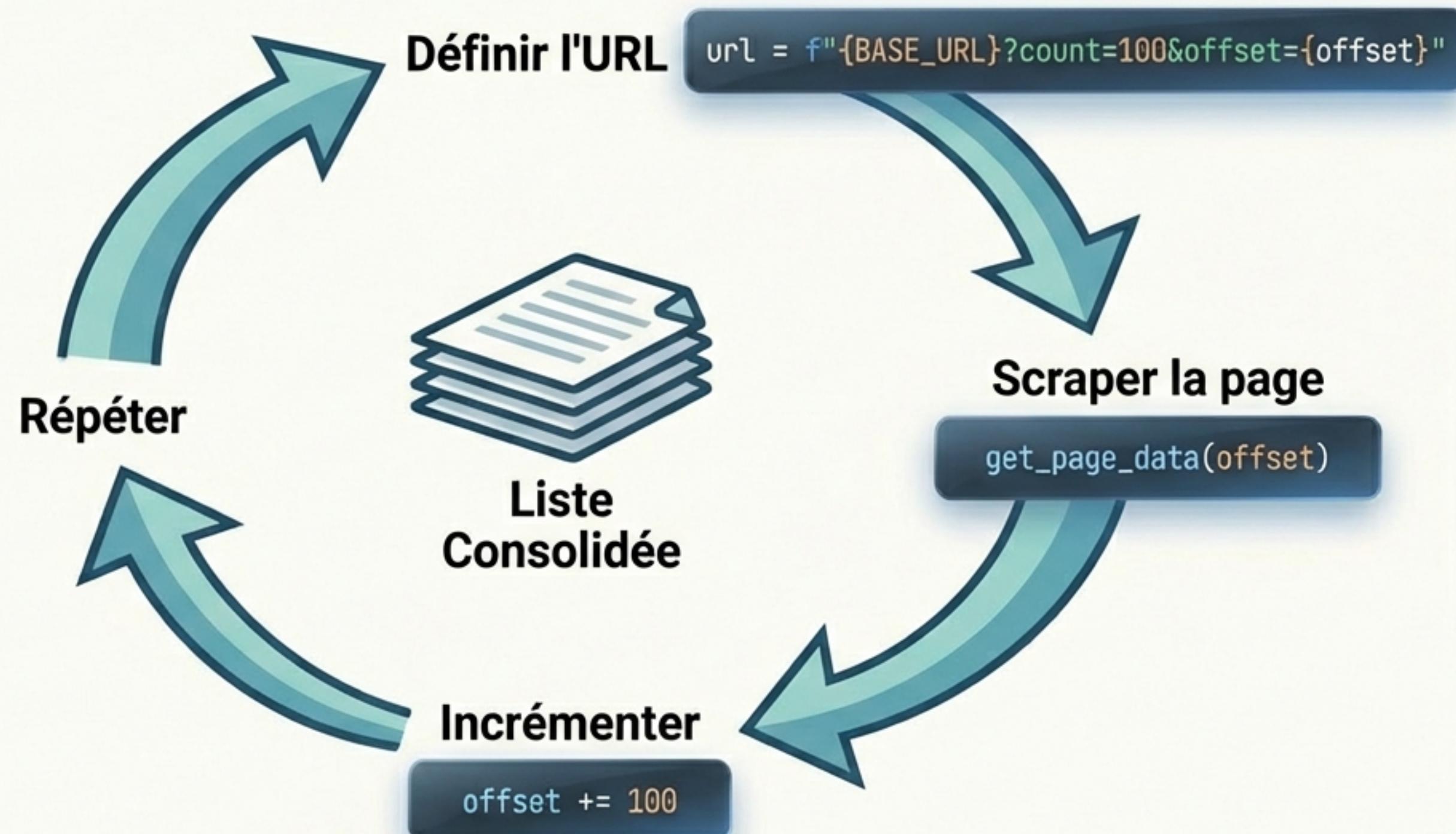
L'Alchimie du Nettoyage : La Fonction `parse_value`

Transformation des abréviations (Texte) en valeurs numériques (Float).



Automatisation à Grande Échelle : La Boucle

Le site n'affiche que 100 résultats par page. Il faut paginer.

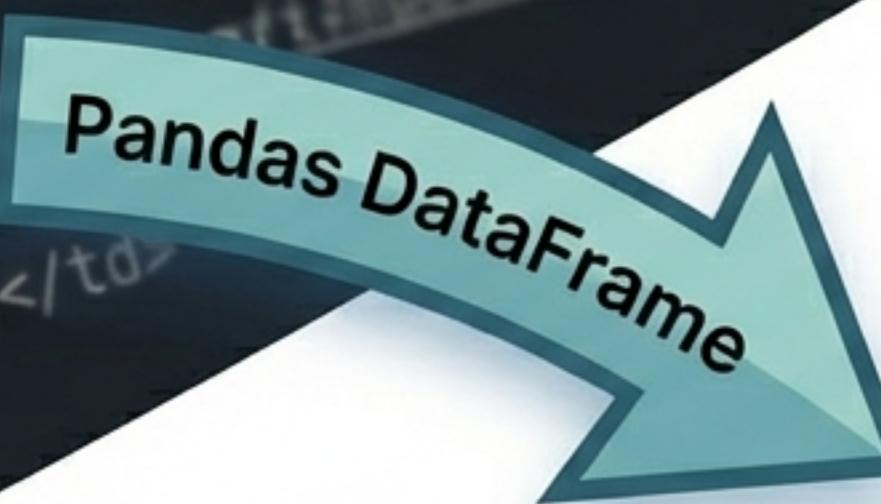


Page 1:
Offset 0
(Top 1-100)

Page 2:
Offset 100
(Top 101-200)

Stockage Structuré : Du Chaos au DataFrame

```
<div>
<div>
<span class="yf-1w0... "></span>
<span class="yf-1w1... "></span>
</div>
<table>
<tr>
<tr>
<td href="#">razino.
<td href="#">NIVN
<td href="#">PLUG
</tr>
</table>
<!-- class="yf-1w1... -->
```



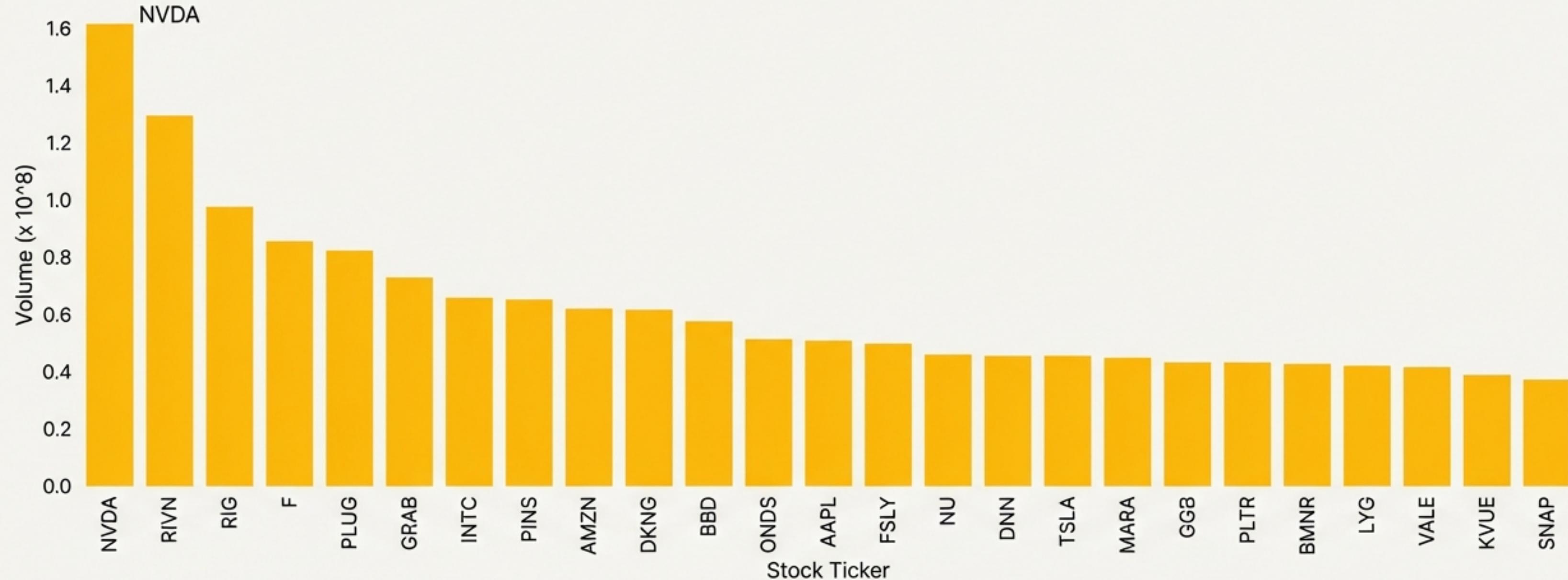
Ticker	Price	Change	Volume
NVDA	182.78	-2.24%	161,888,000
RIVN	17.73	+26.64%	127,927,000
PLUG	1.89	+3.85%	84,314,000

Le Top 5 : Les Rois du Volume

1. NVDA (NVIDIA)	Price \$182.78 	Market Cap \$4.45T
2. RIVN (Rivian)	Price \$17.73  +26.64%	Forte Volatilité
3. RIG (Transocean)	Price \$6.54	Volume 98M
4. F (Ford Motor)	Price \$14.12	
5. PLUG (Plug Power)	Price \$1.89	Penny Stock

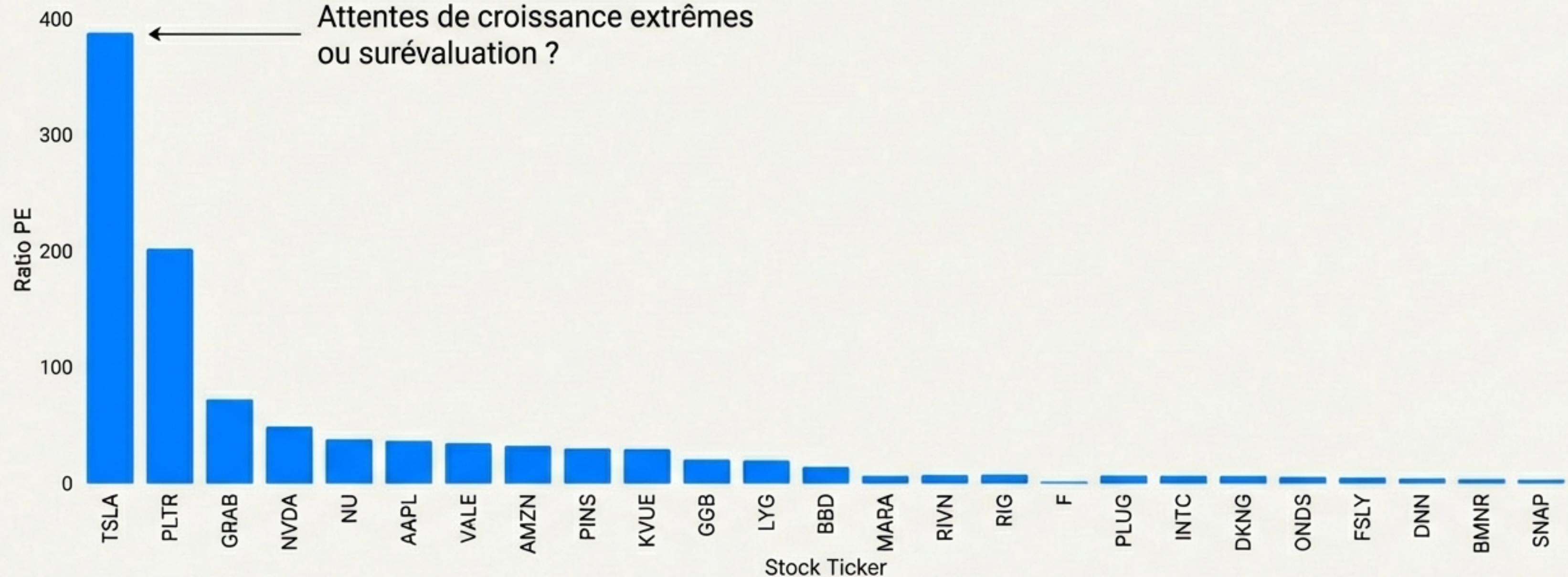
Le code nous permet de trier instantanément des milliers de lignes pour isoler ces opportunités.

Visualisation : Où va l'argent ? (Volume)



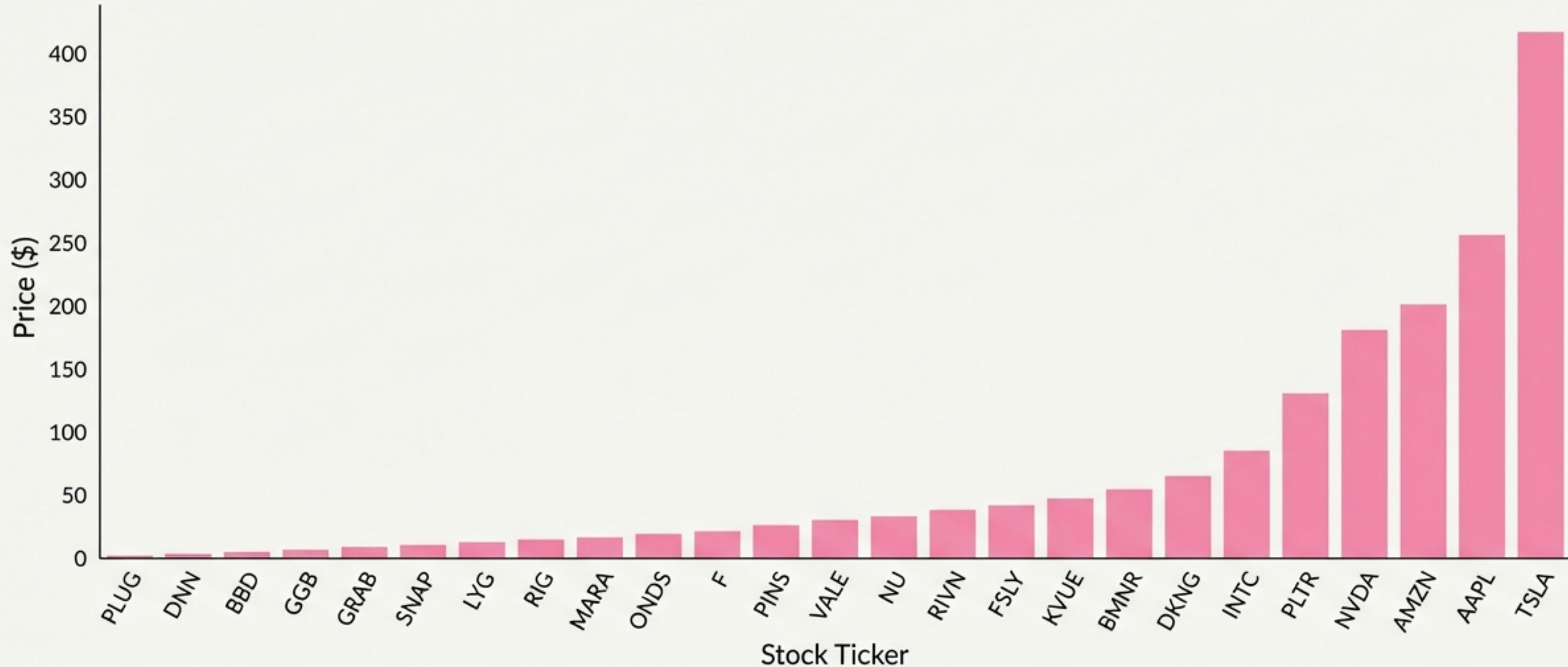
Quelques géants captent la majorité de l'attention du marché.

Visualisation : La Valorisation (Ratio PE)



Quelques géants captent la majorité de l'attention du marché.

Visualisation : Le Coût d'Entrée (Prix)



Disparité totale : des 'Penny stocks' aux géants de la Tech.

Conclusion & Perspectives



Un pipeline autonome : du Web à la Décision.

- Accomplissement : Transformation de données non structurées en tableau d'analyse.
- Futur : Intégration de l'analyse de sentiment (News) et des données historiques.

*Ce projet est un **tremplin** vers une meilleure compréhension des mécanismes **financiers**.*