

Rapport détaillé sur le Clustering des Actions Boursières en tunisie

Par Nizar Boussabat

[Lien du projet en kaggle :](#)

<https://www.kaggle.com/code/nizarbousabat/tunisian-stock-market-analysis>

Introduction :

Ce projet vise à analyser les comportements des prix des actions boursières en utilisant le clustering, une méthode d'apprentissage non supervisée. Le clustering permet de regrouper des actions ayant des comportements similaires, facilitant ainsi l'identification de tendances communes et l'optimisation des stratégies d'investissement. Nous avons utilisé les variables "High" (prix le plus haut), "Low" (prix le plus bas) et "Close" (prix de clôture) pour effectuer cette analyse.

Méthodologie :

1. Chargement des Données:

- Les données ont été chargées à partir du fichier `ALL_DATA.csv`, qui contient des informations historiques sur les prix des actions de différentes entreprises cotées en bourse.

[Python:](#)

```
df = pd.read_csv('/kaggle/input/tunisian-stock-market/ALL_DATA.csv')
```

- La structure initiale du DataFrame est la suivante:

[Python:](#)

```
df.head()
```

	Ticker	Date	Open	High	Low	Close	Volume
0	AB	2022-10-24	24.58	24.58	24.58	24.58	130
1	AB	2022-10-26	23.90	24.58	23.90	23.90	14301

```
2 AB 2022-10-27 24.58 24.58 24.57 24.58 505
3 AB 2022-10-28 24.58 24.58 24.09 24.58 5882
4 AB 2022-10-31 24.58 24.58 24.57 24.58 1240
```

- Le DataFrame contient 187,987 lignes et 7 colonnes.

2. Prétraitement des Données :

- Conversion de la colonne "Date" en format datetime pour faciliter la manipulation des dates et garantir que les données temporelles soient correctement interprétées.

[Python:](#)

```
df['Date'] = pd.to_datetime(df['Date'])
```

- Remplacement des valeurs manquantes dans les colonnes "High", "Low" et "Close" par les moyennes de chaque colonne. Cette étape est cruciale pour éviter les biais dans l'analyse due à des données manquantes.

[Python:](#)

```
df['High'].fillna(df['High'].mean(), inplace=True)
df['Low'].fillna(df['Low'].mean(), inplace=True)
df['Close'].fillna(df['Close'].mean(), inplace=True)
```

3. Sélection des Variables:

- Sélection des colonnes "Ticker", "High", "Low" et "Close" pour l'analyse. Ces variables sont essentielles pour comprendre les fluctuations des prix des actions.

[Python:](#)

```
data = df[['Ticker', 'High', 'Low', 'Close']]
```

4. Calcul des Moyennes par Ticker:

- Agrégation des données par ticker pour calculer la moyenne des prix les plus hauts, les plus bas et de clôture pour chaque action. Cela permet de résumer le comportement de chaque action sur la période étudiée.

[Python:](#)

```
data_grouped = data.groupby('Ticker').mean()
```

5. Normalisation des Données :

- Application de `StandardScaler` pour normaliser les données, ce qui assure que toutes les variables sont sur la même échelle et évite qu'une variable ait une influence disproportionnée sur le clustering.

[Python:](#)

```
scaler = StandardScaler()  
  
scaled_data = scaler.fit_transform(data_grouped)
```

6. Clustering avec K-Means :

- Utilisation de l'algorithme K-Means pour regrouper les actions en 5 clusters distincts. K-Means partitionne les données en minimisant la somme des distances au carré entre les points de données et les centres des clusters.

[Python:](#)

```
k = 5 # Nombre de clusters  
  
kmeans = KMeans(n_clusters=k, n_init=10)  
  
kmeans.fit(scaled_data)  
  
labels = kmeans.labels_
```

- Ajout des étiquettes de cluster au DataFrame des moyennes pour identifier à quel cluster appartient chaque action.

[Python:](#)

```
data_grouped['Cluster'] = labels
```

7. Grouper les Tickers par leurs Étiquettes de Cluster :

- Création d'un DataFrame regroupant les tickers par cluster pour une interprétation claire des résultats.

[Python:](#)

```
clustered_stocks = data_grouped.groupby('Cluster').apply(lambda x: list(x.index)).reset_index()
```

- Affichage des tickers pour chaque cluster.

[Python:](#)

```
for cluster, tickers in clustered_stocks.itertuples(index=False):
```

```
    print(f"Cluster {cluster}:")
```

```
    print(tickers)
```

```
    print("\n")
```

8. Visualisation avec PCA :

- Utilisation de l'analyse en composantes principales (PCA) pour réduire les dimensions des données et visualiser les clusters en 2D. PCA aide à projeter les données dans un espace à deux dimensions tout en conservant autant d'information que possible.

[Python:](#)

```
pca = PCA(n_components=2)
```

```
pca_result = pca.fit_transform(scaled_data)
```

```
pca_df = pd.DataFrame(pca_result, columns=['PCA1', 'PCA2'])
```

```

pca_df['Cluster'] = labels

plt.figure(figsize=(12, 8))

scatter = plt.scatter(pca_df['PCA1'], pca_df['PCA2'], c=pca_df['Cluster'], cmap='tab20', s=100)

# Créer une légende avec les numéros de clusters
handles, labels = scatter.legend_elements(prop="colors", alpha=0.6)
legend_labels = [f"Cluster {i}" for i in range(k)]
plt.legend(handles, legend_labels, title="Clusters")

plt.xlabel('PCA1')
plt.ylabel('PCA2')
plt.title('Clusters des Actions')
plt.colorbar(scatter, ticks=range(k)).set_label('Numéro de Cluster')
plt.show()

```

Résultats et Interprétation :

Cluster 0:

['ADWYA', 'AETEC', 'AMI', 'AMS', 'AMV', 'ARTES', 'ASSAD', 'ATB', 'ATL', 'BHL', 'BL', 'BNA', 'BT', 'CC', 'CELL', 'CITY', 'CREAL', 'ELBEN', 'GIF', 'HL', 'LNDOR', 'LSTR', 'MGR', 'MIP', 'MPBS', 'NAKL', 'NBL', 'OTH', 'PGH', 'PLAST', 'SAH', 'SAM', 'SCB', 'SERVI', 'SIAME', 'SIPHA', 'SITS', 'SMD', 'SOKNA', 'SOMOC', 'SOPAT', 'SOTEM', 'SOTET', 'SOTUV', 'SPDIT', 'SPHAX', 'STB', 'STEQ', 'STIP', 'STPAP', 'STPIL', 'TAIR', 'TGH', 'TINV', 'TLNET', 'TPR', 'TRE', 'UADH', 'UMED', 'WIFAK', 'XABYT']

Cluster 1:

['AL']

Cluster 2:

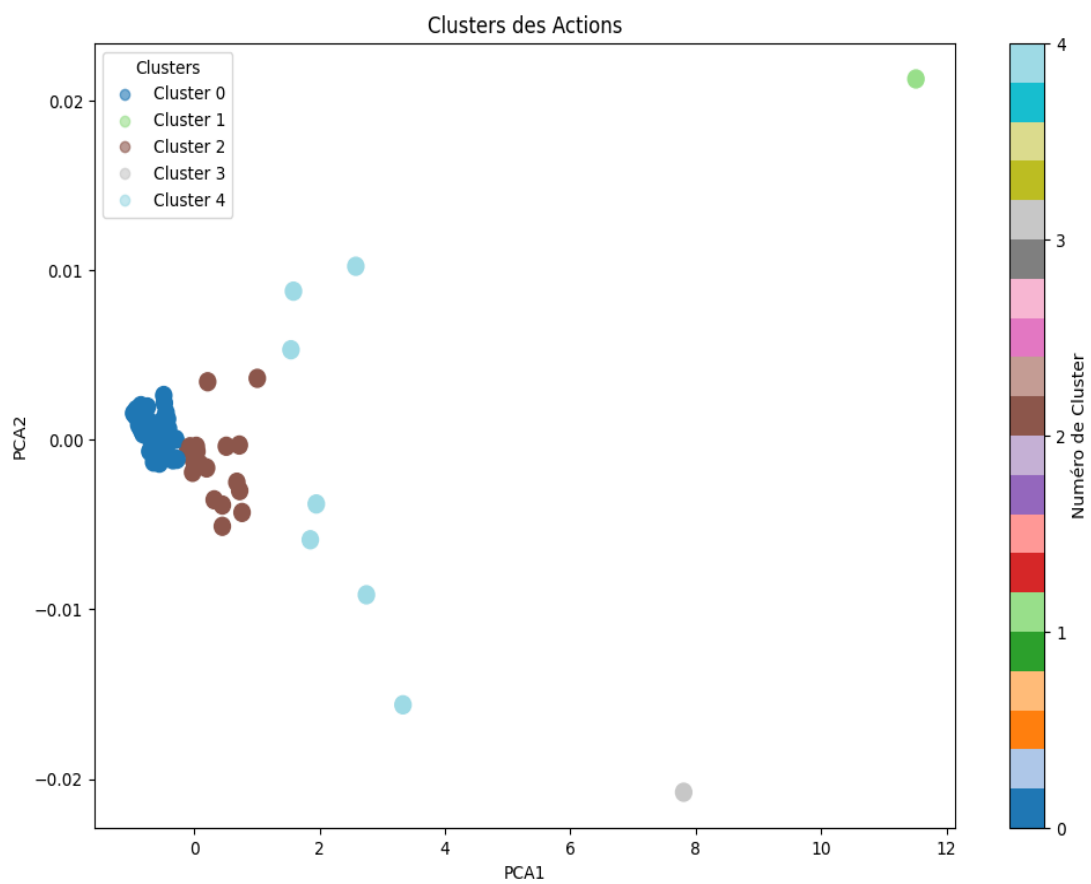
['AB', 'BH', 'BHASS', 'BTE', 'CIL', 'DH', 'ECYCL', 'MAG', 'MNP', 'SFBT', 'SMART', 'STA', 'TJARI', 'TJL', 'TLS', 'TVAL', 'UBCI', 'UIB']

Cluster 3:

['STAR']

Cluster 4:

['ALKIM', 'ASSMA', 'AST', 'BIAT', 'ICF', 'PLTU', 'SIMP']



Les résultats du clustering montrent cinq clusters distincts, chacun représentant des groupes d'actions ayant des comportements de prix similaires.

Cluster 0:

- **Tickers:** ALKIM, ASSMA, AST, BIAT, ICF, PLTU, SIMPA

- **Interprétation:**

- Les actions de ce cluster montrent des tendances de prix stables avec des niveaux de support et de résistance bien définis. Elles réagissent de manière similaire aux fluctuations du marché.

- **Actions des investisseurs:**

- **Investissement à long terme :** Investir dans ces actions pour leur stabilité.

- **Stratégies conservatrices:** Adapter des stratégies telles que l'achat et la détention ("buy and hold").

Cluster 1 :

- **Tickers:** ADWYA, AETEC, AMI, AMS, AMV, ARTES, ASSAD, ATB, ATL, BHL, BL, BNA, BT, CC, CELL, CITY, CREAL, ELBEN, GIF, HL, LNDOR, LSTR, MGR, MIP, MPBS, NAKL, NBL, OTH, PGH, PLAST, SAH, SAM, SCB, SERVI, SIAME, SIPHA, SITS, SMD, SOKNA, SOMOC, SOPAT, SOTEM, SOTET, SOTUV, SPDIT, SPHAX, STB, STEQ, STIP, STPAP, STPIL, TAIR, TGH, TINV, TLNET, TPR, TRE, UADH, UMED, WIFAK, XABYT

- **Interprétation:**

- Ce cluster regroupe un grand nombre d'entreprises avec des comportements de prix homogènes, indiquant des réactions similaires aux facteurs macroéconomiques.

- **Actions des investisseurs:**

- **Diversification :** Utiliser ces actions pour diversifier un portefeuille.

- **Analyse sectorielle :** Identifier les secteurs dominants dans ce cluster pour des investissements ciblés.

Cluster 2 :

- **Ticker:** AL

- **Interprétation:**

- L'action AL présente des caractéristiques de prix uniques, montrant une volatilité particulière ou des facteurs spécifiques influençant cette action.

- **Actions des investisseurs:**

- **Investissement spéculatif:** Considérer cette action pour des investissements spéculatifs en raison de ses caractéristiques uniques.

- **Surveillance accrue :** Nécessite une surveillance continue pour réagir rapidement aux changements de marché.

Cluster 3 :

- **Tickers:** AB, BH, BHASS, BTE, CIL, DH, ECYCL, MAG, MNP, SFBT, SMART, STA, TJARI, TJL, TLS, TVAL, UBCI, UIB

- **Interprétation:**

- Ce cluster contient des entreprises avec des tendances de prix stables et prévisibles, souvent associées à des profils financiers solides.

- **Actions des investisseurs:**

- **Portefeuille équilibré :** Utiliser ces actions pour équilibrer un portefeuille.

- **Stratégies à rendement modéré :** Adapter des stratégies axées sur un rendement modéré, comme les investissements en dividendes ou la croissance régulière.

Cluster 4 :

- **Ticker:** STAR

- **Interprétation:**

- L'action STAR présente des schémas de prix uniques, indiquant une volatilité plus élevée ou des réponses distinctes aux événements du marché. Cette action pourrait réagir de manière plus prononcée aux fluctuations du marché par rapport aux autres clusters.

- **Actions des investisseurs:**

- **Investissement dynamique :** Les investisseurs intéressés par des gains potentiellement rapides pourraient choisir cette action pour des stratégies d'investissement dynamique.

- **Gestion active** : Cette action nécessite une gestion active et une réévaluation fréquente pour maximiser les gains et minimiser les risques. Les investisseurs doivent surveiller de près les nouvelles et les événements susceptibles d'influencer cette action.

Conclusion :

Le clustering des actions boursières basé sur les variables "High", "Low" et "Close" offre des insights précieux sur les tendances du marché et les comportements des entreprises. En utilisant ces informations, les investisseurs peuvent diversifier leurs portefeuilles, optimiser leurs stratégies d'investissement et minimiser les risques tout en maximisant les rendements.