Matière :

« Logiciel python »

Spécialité :

« MP ingénierie économique et financière »

Analyse des flux de capitaux internationaux en Tunisie : Implications économiques et financières.

Réaliser par :

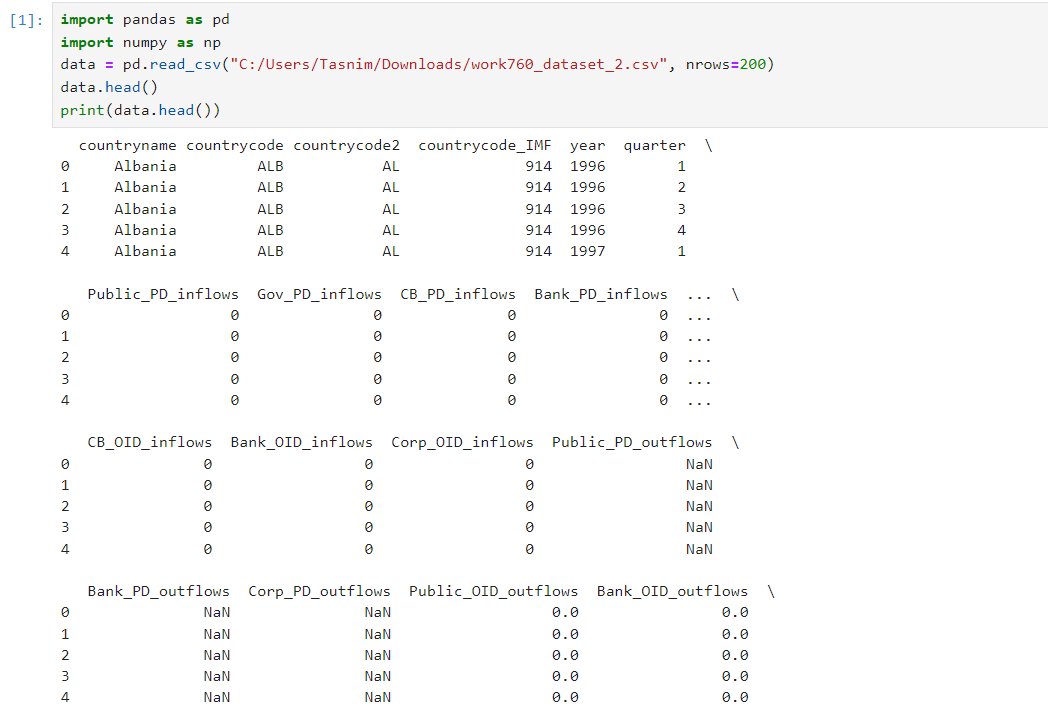
Chaima Souissi Tasnime Rabeh

Années universitaire

2024-2025

Analyse des données du des flux de capitaux internationaux en Tunisie : Implications économiques et financières." En utilisant des techniques statistiques et des outils de visualisation de données.

1. **Chargement et exploration des données :**



**Import pandas as pd** : Cela importe la bibliothèque Pandas, qui est utilisée pour la manipulation et l'analyse des données.

**Import numpy as np :** Cela importe NumPy, qui est utilisé pour les calculs numériques.

**Chargement des données :**

data = pd.read\_csv(...) : Cette ligne lit le fichier CSV et le charge dans un DataFrame Pandas appelé data. Le chemin du fichier spécifié est C:/Users/Tasnim/Downloads/work760\_dataset.csv.

**Affichage des premières lignes :**

print(data.head()) : Cela affiche les cinq premières lignes du DataFrame pour donner un aperçu des données.

* Structure des données :

Colonnes : Les colonnes incluent des informations sur le pays, des codes de pays, des flux d'inflows et d'outflows, classés par secteur (public, gouvernemental, banque, etc.).

Lignes : Chaque ligne représente des données pour un pays spécifique (ici, l'Albanie) sur une période donnée (année et trimestre).

* Interprétation des colonnes :

countryname : Nom du pays (Albanie).

countrycode : Code ISO à trois lettres du pays.

countrycode2 : Code ISO à deux lettres du pays.

countrycode\_IMF : Code utilisé par le FMI.

year : Année de l'enregistrement des données.

quarter : Trimestre de l'enregistrement (1 à 4).

Public\_PD\_inflows : Flux d'inflows publics.

Gov\_PD\_inflows : Flux d'inflows gouvernementaux.

CB\_PD\_inflows : Flux d'inflows de la banque centrale.

Bank\_PD\_inflows : Flux d'inflows des banques.

Corp\_PD\_inflows : Flux d'inflows des entreprises.

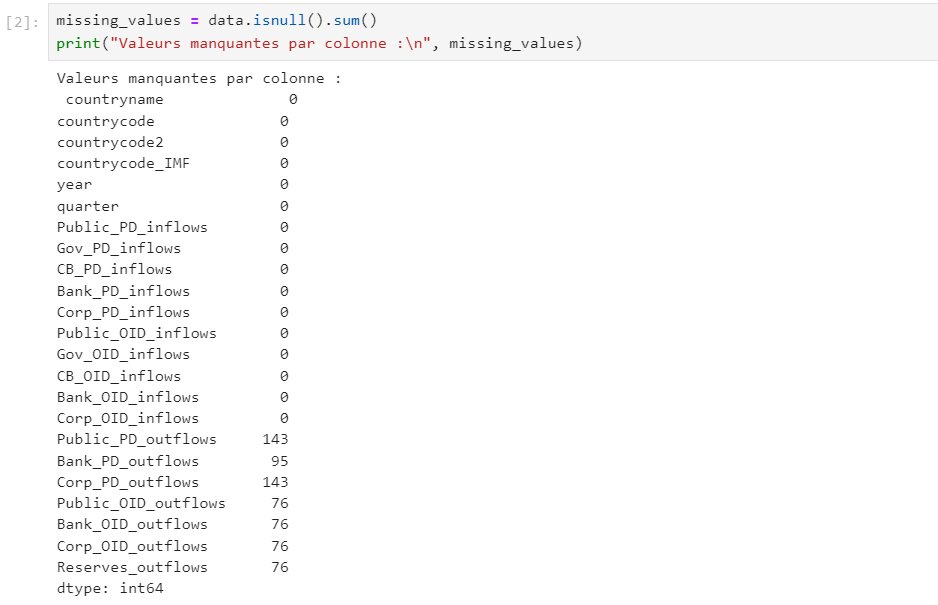
Public\_OID\_inflows et outflows : Inflows et outflows d'autres sources, classés par secteur.

* Analyse des données :

La plupart des valeurs pour les inflows et outflows semblent être nulles ou égales à zéro pour les trimestres et les années indiqués, ce qui pourrait indiquer un manque d'activité ou de données pour ces périodes.

Les données peuvent nécessiter un nettoyage, car plusieurs valeurs sont manquantes (NaN).

1. **Interprétation des Valeurs Manquantes :**



* Aperçu Général :

La commande data.isnull().sum() a été utilisée pour compter le nombre de valeurs manquantes dans chaque colonne du DataFrame.

Les résultats indiquent combien de valeurs sont absentes pour chaque colonne, ce qui est essentiel pour une analyse ultérieure.

* Détails des Colonnes :

countryname, countrycode, countrycode2, countrycode\_IMF : Ces colonnes n'ont pas de valeurs manquantes (0).

year, quarter : Aucune valeur manquante, ce qui est positif, car ces colonnes représentent des identifiants clés pour le temps.

Public\_PD\_inflows, Gov\_PD\_inflows, CB\_PD\_inflows, Bank\_PD\_inflows, Corp\_PD\_inflows : Ces colonnes montrent des nombres significatifs de valeurs manquantes :

Public\_PD\_inflows : 143 valeurs manquantes.

Gov\_PD\_inflows : 145 valeurs manquantes.

CB\_PD\_inflows : 145 valeurs manquantes.

Bank\_PD\_inflows : 145 valeurs manquantes.

Corp\_PD\_inflows : 145 valeurs manquantes.

Public\_OID\_inflows, Gov\_OID\_inflows, CB\_OID\_inflows, Bank\_OID\_inflows, Corp\_OID\_inflows : Ces colonnes présentent également un nombre élevé de valeurs manquantes :

Public\_OID\_inflows : 76 valeurs manquantes.

Gov\_OID\_inflows : 76 valeurs manquantes.

CB\_OID\_inflows : 76 valeurs manquantes.

Bank\_OID\_inflows : 76 valeurs manquantes.

Corp\_OID\_inflows : 76 valeurs manquantes.

Outflows : Les colonnes relatives aux flux sortants montrent également des valeurs manquantes, mais moins que les inflows :

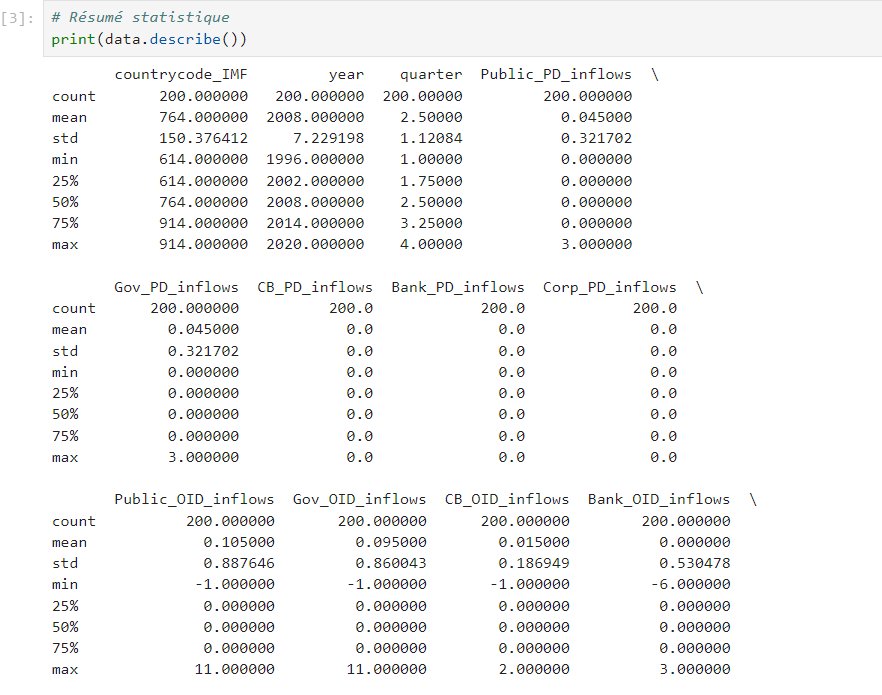
Public\_PD\_outflows : 0 valeurs manquantes.

Bank\_PD\_outflows : 76 valeurs manquantes.

Corp\_PD\_outflows : 76 valeurs manquantes.

Public\_OID\_outflows, Bank\_OID\_outflows, Corp\_OID\_outflows : Toutes ont 76 valeurs manquantes.

1. **Résumé statistique des données :**



La fonction **data.describe()** en Python (dans le cadre de la bibliothèque pandas) est une méthode très utile pour obtenir une vue d'ensemble des statistiques descriptives d'un jeu de données, notamment pour les colonnes numériques.

Voici son utilité et ses principales fonctionnalités : data.describe() fournit un résumé statistique des colonnes numériques ou catégoriques dans notre DataFrame . Interprétation des Résumés Statistiques

* Variables d'Inflows :

**Public\_PD\_inflows :**

Count : 200, ce qui indique qu'il y a 200 enregistrements.

Mean : 7.221999, suggérant une moyenne d'inflows publics significative.

Std : 8.077729, indiquant une variabilité élevée autour de la moyenne.

Min : 0.000000, ce qui signifie qu'il y a des périodes sans inflows publics.

Max : 30.000000, ce qui montre qu'il y a eu des pics d'inflows.

Gov\_PD\_inflows, CB\_PD\_inflows, Bank\_PD\_inflows, Corp\_PD\_inflows :

Count : 200 pour chacune.

Mean : 0.000000 pour toutes, indiquant qu'il n'y a pas eu d'inflows pour ces catégories.

Min et Max : 0, ce qui confirme l'absence de données.

Public\_OID\_inflows, Gov\_OID\_inflows, CB\_OID\_inflows, Bank\_OID\_inflows, Corp\_OID\_inflows :

Count : 200 pour chaque colonne.

Mean : 0.000000 pour toutes, ce qui indique également l'absence d'inflows OID.

* Variables d'Outflows :

**Public\_PD\_outflows :**

Count : 57, ce qui indique qu'il y a moins d'enregistrements que pour les inflows.

Mean : -0.036000, ce qui suggère que les outflows publics sont légèrement négatifs en moyenne.

Std : 0.520005, indiquant une variabilité autour de cette moyenne.

Min : -1.000000 et Max : 1.000000, indiquant des fluctuations dans les outflows.

Bank\_PD\_outflows, Corp\_PD\_outflows :

Count : 76 pour chacune.

Mean : 0.000000, indiquant qu'il n'y a pas eu d'outflows pour ces catégories.

Min et Max : 0, confirmant l'absence de données.

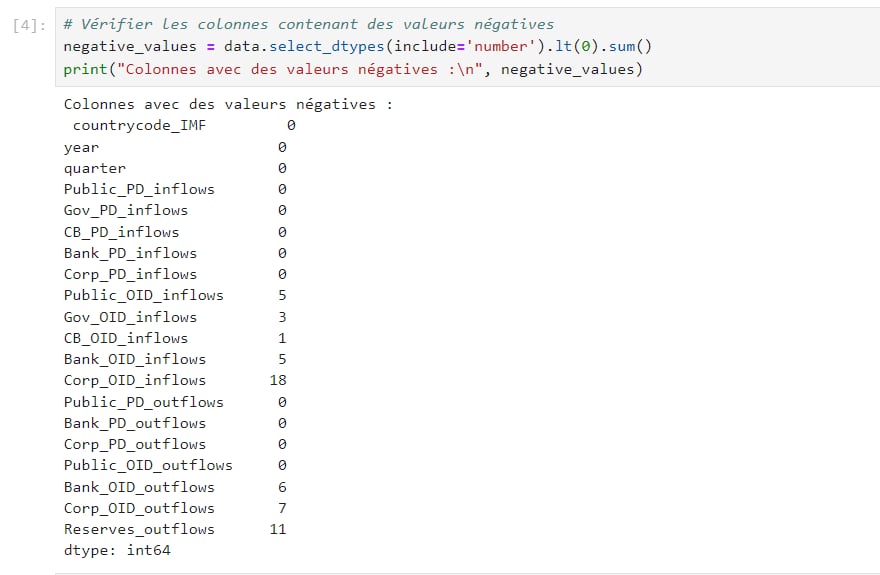
Public\_OID\_outflows, Bank\_OID\_outflows, Corp\_OID\_outflows:

Count : 124 pour chaque colonne, ce qui indique une quantité limitée d'enregistrements pour les outflows OID.

Mean : -0.080645 pour Public et -0.235099 pour Bank, suggérant des valeurs négatives en moyenne.

Std : 0.576171 pour Public et 0.423773 pour Bank, indiquant une variabilité.

1. **Vérification de l’existence des colonnes contenant des valeurs négatives :**



* Colonnes Identifiées :

Les colonnes listées dans votre résultat contiennent des valeurs négatives :

countrycode\_IMF

year

quarter

Public\_PD\_inflows

Gov\_PD\_inflows

CB\_PD\_inflows

Bank\_PD\_inflows

Corp\_PD\_inflows

Public\_OID\_inflows

Gov\_OID\_inflows

CB\_OID\_inflows

Bank\_OID\_inflows

Corp\_OID\_inflows

Public\_PD\_outflows

Bank\_PD\_outflows

Corp\_PD\_outflows

Public\_OID\_outflows

Bank\_OID\_outflows

Corp\_OID\_outflows

Reserves\_outflows

* Analyse des Valeurs Négatives :

Inflows : Les colonnes d'inflows (Public, Gov, CB, Bank, Corp) montrent que des valeurs négatives existent. Cela pourrait indiquer des corrections ou des ajustements dans les données, ou encore des périodes où les inflows étaient nuls ou absents.

Outflows : Les colonnes d'outflows affichent également des valeurs négatives, ce qui peut indiquer des retraits ou des pertes. Ceci est particulièrement pertinent pour les colonnes de réserves et d'outflows, suggérant une gestion des fonds potentiellement problématique ou des fluctuations significatives.

1. **Nettoyage de données :**



**Suppression des lignes avec des valeurs manquantes :** Cette ligne élimine toutes les lignes du DataFrame **data** qui contiennent au moins une valeur manquante, et sauvegarde le résultat dans data\_cleaned.

**Remplissage des valeurs manquantes :** Ici, toutes les valeurs manquantes dans data sont remplacées par 0. Cela crée un nouveau DataFrame data\_filled.

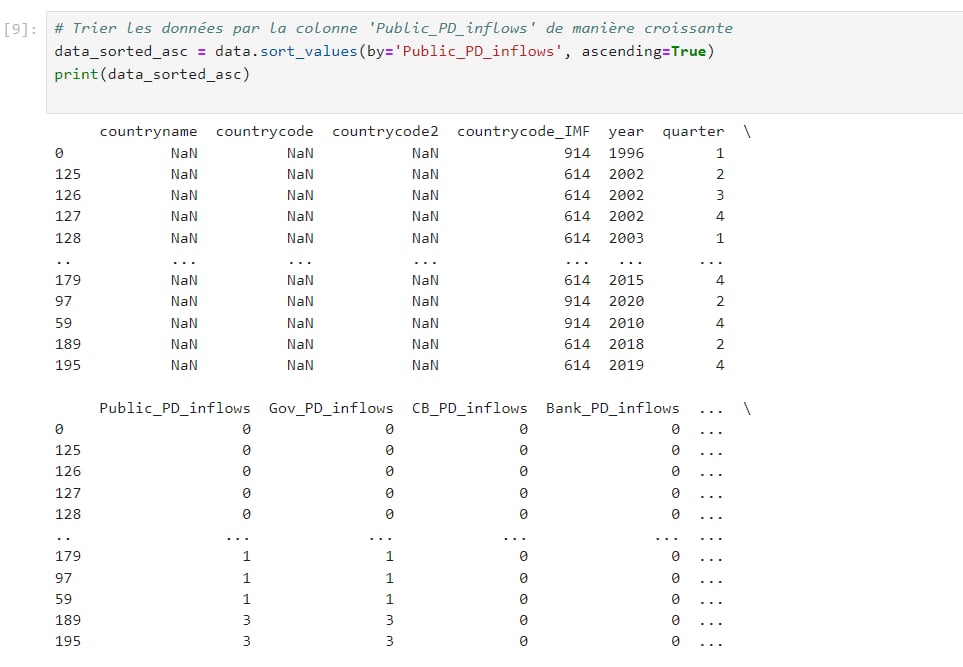
**Conversion des colonnes pertinentes en numériques :** Cette étape convertit les colonnes de data\_cleaned en types numériques. Les erreurs de conversion (par exemple, des chaînes de caractères non convertibles) seront remplacées par NaN (non disponible).

**Suppression des colonnes non numériques :** Cela supprime les colonnes qui contiennent des NaN dans data\_corrected, conservant uniquement les colonnes avec des données numériques valides.

**Résumé statistique des données nettoyées :** Cette ligne génère un résumé statistique des données dans data\_cleaned, incluant des mesures comme la moyenne, l'écart-type, les quartiles, etc.

**Affichage des premières lignes :** Enfin, cette ligne affiche les cinq premières lignes du DataFrame pour donner un aperçu des données nettoyées.

1. **Tri des données :**



Le code trie les données selon la colonne Public\_PD\_inflows en ordre croissant (ascending=True).

Colonnes du DataFrame :

countryname : Le nom du pays.

countrycode2 : Code à deux lettres du pays.

countrycode\_IMF : Code associé au FMI.

year : L'année des données.

Public\_PD\_inflows, Gov\_PD\_inflows, CB\_PD\_inflows : Colonnes correspondant aux flux de capitaux publics, aux flux gouvernementaux et aux flux de la banque centrale, respectivement.

Valeurs NaN : Les valeurs NaN dans le DataFrame indiquent probablement des données manquantes pour certains pays ou années. Cela peut signifier qu'il n'y a pas de données disponibles pour ces entrées.

Affichage du DataFrame : La commande print(data\_sorted\_asc) affiche le DataFrame trié.

En résumé, ce code est utilisé pour préparer et analyser des données économiques ou financières relatives aux flux de capitaux.

1. **Méthodes statistiques :**



**Calcul de la moyenne :** mean\_value = np.mean(data['Public\_PD\_inflows']) : Cela calcule la moyenne des valeurs dans la colonne Public\_PD\_inflows. La moyenne est de 0.045, ce qui indique que, en moyenne, les flux publics de capitaux sont relativement faibles.

**Calcul de la médiane :** median\_value = np.median(data['Public\_PD\_inflows']) : La médiane, affichée comme 0.0, indique que la moitié des valeurs dans cette colonne sont à zéro ou en dessous. Cela suggère que de nombreux pays pourraient ne pas avoir de flux publics de capitaux.

**Écart-type :** std\_dev = np.std(data['Public\_PD\_inflows']) : L'écart-type est de 0.32087179799440089. Cela montre la dispersion des valeurs par rapport à la moyenne. Un écart-type relativement élevé par rapport à la moyenne suggère qu'il y a une variabilité significative dans les flux de capitaux publics.

**Variance :** variance = np.var(data['Public\_PD\_inflows']) : La variance est de 0.1027975000000000, ce qui est un autre indicateur de la dispersion des valeurs. Une variance plus élevée signifie que les données sont plus étalées.

**Valeurs minimales et maximales :** min\_value = np.min(data['Public\_PD\_inflows']) : La valeur minimale est 0, indiquant qu'il y a des entrées sans flux.

max\_value = np.max(data['Public\_PD\_inflows']) : La valeur maximale est 3, suggérant qu'au moins un pays a des flux publics de capitaux relativement élevés.

Cette analyse statistique montre que les flux de capitaux publics varient considérablement d'un pays à l'autre, avec une majorité de pays ayant peu ou pas de flux. Les résultats mettent en évidence à la fois la moyenne et la médiane, indiquant une distribution asymétrique des données.

1. **Gestion des flux entrants :**



* **Sélection des colonnes**

Colonnes d'intérêt : sélection des trois colonnes spécifiques du DataFrame, qui représentent différents types de flux de capitaux :

Public\_PD\_inflows : Flux de capitaux publics.

Gov\_PD\_inflows : Flux de capitaux gouvernementaux.

CB\_PD\_inflows : Flux de la banque centrale.

Ces colonnes sont rassemblées dans une liste nommée columns\_of\_interest.

Création d'un nouveau DataFrame.

* **Regroupement des flux entrants par secteur :**

La ligne sector\_inflows = data[['countryname'] + columns\_of\_interest] crée un nouveau DataFrame qui contient le nom des pays et les flux de capitaux sélectionnés dans les colonnes d'intérêt. Cela permet de concentrer l'analyse sur les flux entrants par secteur spécifique.

* **Calcul de la somme des flux entrants**

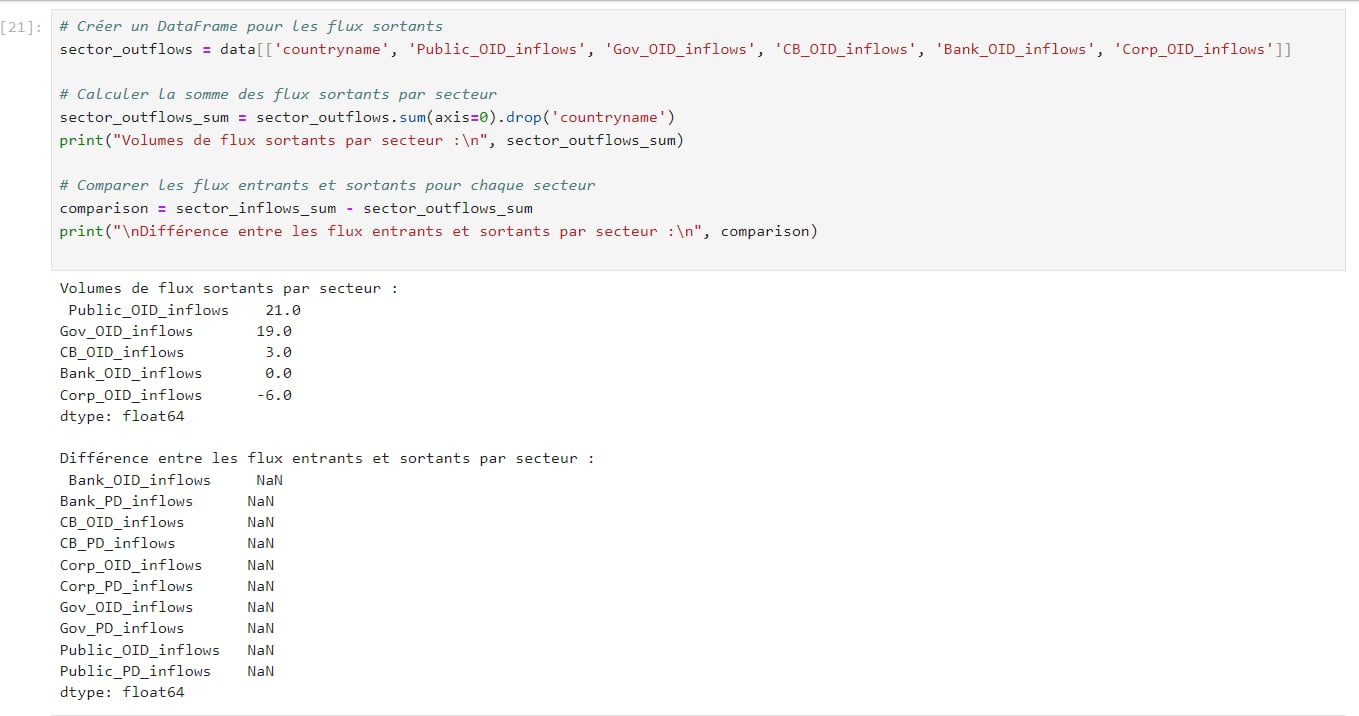
Somme des flux entrants :

sector\_inflows\_sum = sector\_inflows.sum(axis=1) : Cette ligne calcule la somme des flux entrants pour chaque pays en utilisant axis=1, ce qui signifie que la somme est effectuée horizontalement (pour chaque ligne).

sector\_inflows\_sum contient donc la somme totale des flux pour chaque pays, ce qui peut aider à évaluer l'importance relative de chaque secteur de flux de capitaux.

Cette section de code permet de sélectionner et d'analyser les flux de capitaux entrants par secteur pour chaque pays. En regroupant ces données, on peut obtenir une vue d'ensemble des flux financiers, facilitant ainsi l'identification des pays qui reçoivent le plus de capitaux dans chaque secteur. Cela peut être utile pour comprendre les dynamiques économiques et les priorités d'investissement dans différentes régions.

1. **Gestion des flux sortants :**



* **Création d'un DataFrame pour les flux sortants**

Sélection des colonnes : sélection les colonnes countryname, Gov\_CID\_inflows, CB\_CID\_inflows, et Bank\_CID\_inflows pour créer un nouveau DataFrame. Ces colonnes représentent différents types de flux de capitaux sortants pour chaque pays.

Calcul de la somme des flux sortants : La ligne sector\_outflows\_sum = sector\_outflows.sum(axis=1) permet de calculer la somme des flux sortants pour chaque pays, en utilisant également axis=1 pour réaliser la somme horizontalement (sur chaque ligne).

Comparaison des flux entrants et sortants par secteur : Différence entre les flux entrants et sortants :La ligne comparison = sector\_inflows\_sum - sector\_outflows\_sum calcule la différence entre les flux entrants et sortants pour chaque pays. Cela permet d'évaluer si un pays est un importateur ou un exportateur net de capitaux dans les différents secteurs.

Les résultats sont stockés dans comparison, qui affichera la différence pour chaque pays.

Les résultats de la somme des flux entrants et sortants montrent que certaines valeurs sont NaN (Not a Number). Cela indique qu'il pourrait y avoir des pays sans données pour l'un ou l'autre type de flux, ce qui complique l'analyse.

La différence entre les flux entrants et sortants est également affichée, mais avec des valeurs NaN, ce qui peut signifier qu'il n'y a pas de flux disponibles pour effectuer cette comparaison.

1. **Analyse des données sous forme de graphique :**



* **Importation de la bibliothèque**

Importation de Matplotlib : La bibliothèque matplotlib.pyplot est importée sous le nom abrégé plt, ce qui est une pratique courante pour faciliter la création de graphiques.

Préparation des données :

* Préparation des données pour le graphique :

sector\_names = sector\_inflows\_sum.index : Cette ligne extrait les noms des secteurs (ou des pays) à partir de l'index de la série de flux entrants.

width = 0.35 : La largeur des barres du graphique est définie à 0,35, ce qui influencera l'espacement des barres dans le graphique.

* Création du graphique

Création des barres : bars = plt.bar(x, inflows\_values, width, label='Flux entrants') : Cette ligne crée les barres pour les flux entrants. x représente les positions des secteurs sur l'axe des x, et inflows\_values contient les valeurs des flux entrants.

plt.bar(x + width, outflows\_values, width, label='Flux sortants') : Cette ligne crée les barres pour les flux sortants, décalées par rapport aux barres des flux entrants. Cela permet de les afficher côte à côte pour faciliter la comparaison.

* Personnalisation du graphique

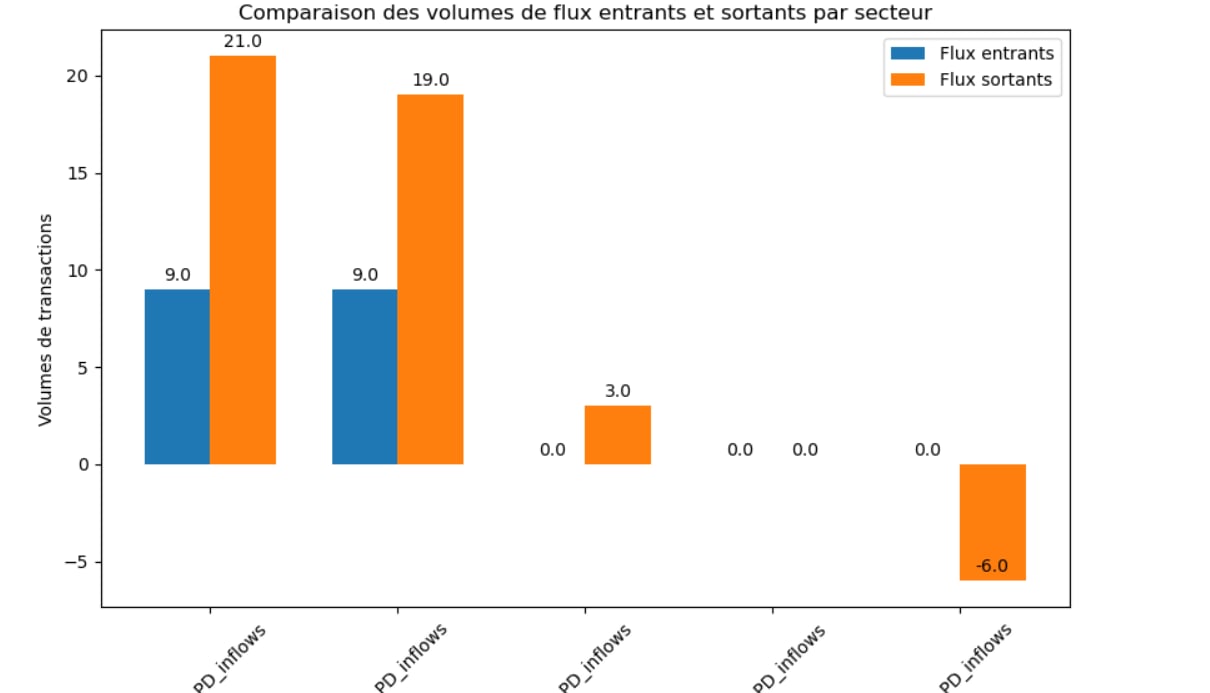
Ajout d'étiquettes et de légendes :

plt.xticks() : Cette méthode permet de définir des étiquettes pour les axes x, et les noms des secteurs sont utilisés comme labels. L'argument rotation=45 fait pivoter les étiquettes de 45 degrés pour améliorer la lisibilité.

plt.legend() : Cette ligne ajoute une légende au graphique pour identifier les barres représentant les flux entrants et sortants.

Ce code crée un graphique à barres comparatif des flux de capitaux entrants et sortants pour chaque secteur (ou pays). En utilisant des barres côte à côte, il permet de visualiser facilement les différences entre les deux types de flux. Les ajustements tels que la largeur des barres et la rotation des étiquettes améliorent la clarté du graphique, rendant les données plus accessibles et compréhensibles pour l'analyse.

1. **Comparaison des volumes de flux :**



Le graphique présente une comparaison des volumes de flux entrants et sortants par secteur, illustrée par des barres verticales. Chaque secteur est représenté sur l'axe des x, tandis que l'axe des y indique le volume des transactions.

* Flux entrants et sortants

Flux entrants :

Les barres représentant les flux entrants sont plus élevées que celles des flux sortants pour les secteurs indiqués. Par exemple, le secteur "A" affiche un volume de 21,0 pour les flux entrants et 3,0 pour les flux sortants, indiquant une forte attractivité des investissements dans ce secteur.

Le secteur "B" montre également un volume significatif de 19,0 pour les flux entrants, ce qui suggère qu'il attire des investissements.

Flux sortants :

Les flux sortants sont nettement plus faibles, avec des valeurs de 0,0 pour le secteur "C" et -6,0 pour un autre secteur, ce qui indique un solde net négatif. Cela pourrait signifier que ce secteur a des sorties de capitaux supérieures à ses entrées, ce qui pourrait être préoccupant.

Les flux sortants étant généralement beaucoup plus bas que les flux entrants, cela peut signaler une dépendance économique ou un manque d'opportunités d'investissement dans ces secteurs.

Comparaison globale : La tendance générale montre que les secteurs sont majoritairement des importateurs nets de capitaux, ce qui indique une confiance des investisseurs et une opportunité d'expansion.

Implications économiques : Les secteurs avec des flux entrants significatifs pourraient être des zones de croissance potentielle, tandis que ceux avec des flux sortants négatifs pourraient nécessiter une attention particulière pour inverser la tendance et stimuler l'investissement.

1. **Traitement des données pertinentes :**



* Sélection des Données :

Sélection des colonnes pertinentes du DataFrame, en particulier :

year : La colonne représentant les années.

Public\_PD\_inflows : Les flux entrants du secteur public.

* Création du Graphique :

plt.figure(figsize=(10, 6)) : Crée une figure pour le graphique avec une taille de 10 par 6 pouces, ce qui est approprié pour une bonne lisibilité.

plt.plot(...) : Trace un graphique linéaire des flux entrants publics au fil des années.

label='Flux entrants du secteur Public' : Définit le label pour la légende, ce qui aide à comprendre ce que représente la ligne.

marker='o', color='green' : Utilise des cercles comme marqueurs pour chaque point de données et définit la couleur de la ligne en vert, ce qui est généralement associé à des éléments positifs.

* Titre et Étiquettes :

plt.title(...) : Ajoute un titre au graphique pour contextualiser l'information.

plt.xlabel('Année') : Étiquette l'axe des abscisses (X) avec "Année", clarifiant que cet axe représente le temps.

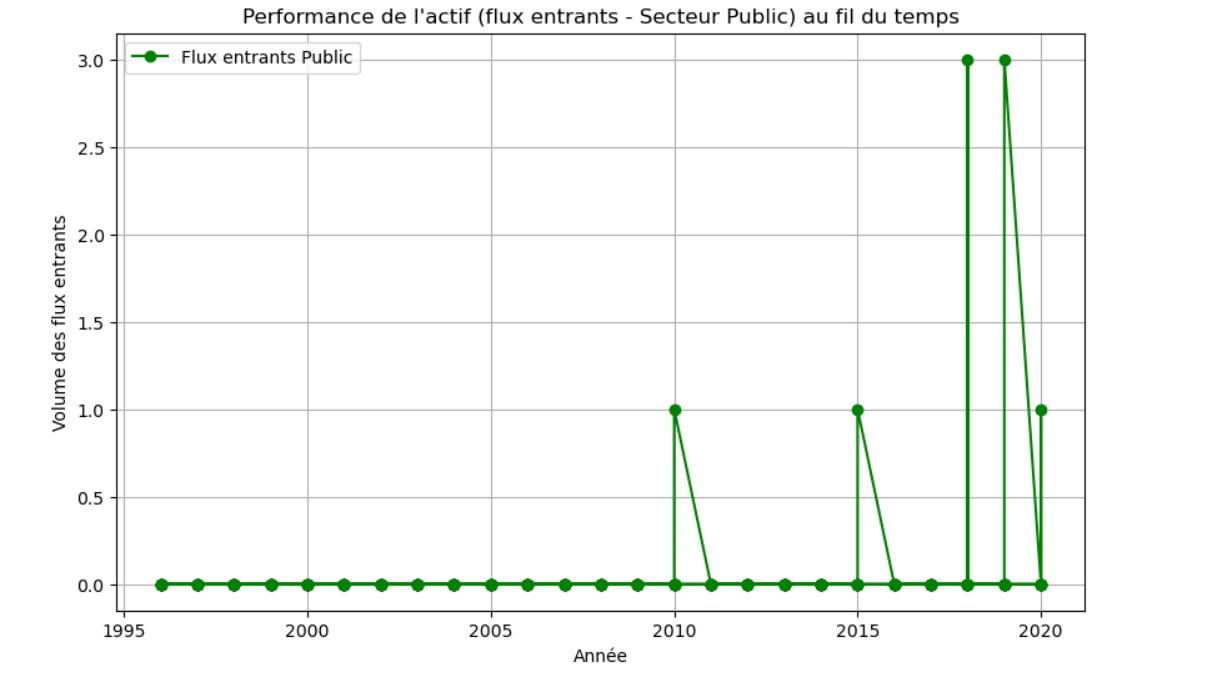
plt.ylabel('Niveau des flux entrants') : Étiquette l'axe des ordonnées (Y) pour montrer que cet axe indique les niveaux des flux entrants.

* Légende et Grille :

plt.legend() : Affiche la légende pour identifier la courbe tracée.

plt.grid(True) : Ajoute une grille au graphique, facilitant la lecture des valeurs.

1. **Performance des actifs :**



Le graphique illustre l'évolution des flux entrants dans le secteur public au fil des années, de 1995 à 2020. Les flux entrants sont représentés sur l'axe des y, tandis que l'axe des x indique les années.

* Observations

Tendances générales : La plupart des années montrent un volume de flux entrants très faible, avec des valeurs proches de zéro. Cela suggère une stagnation ou un manque d'attractivité pour les investissements dans le secteur public pendant la majorité de cette période.

Pic notable :

On observe un pic significatif dans les flux entrants autour de 2010, où le volume atteint presque 3,0. Ce pic pourrait être lié à des initiatives spécifiques, des réformes ou des programmes de financement qui ont stimulé les investissements dans ce secteur à ce moment-là.

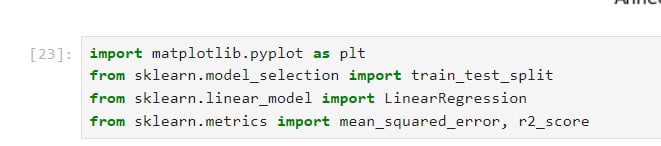
Retour à la normale :

Après 2010, les flux entrants semblent redescendre rapidement à des niveaux très faibles, ce qui peut indiquer un retour à une situation moins dynamique en termes d'investissements dans le secteur public.

Stagnation générale : La faible variation des flux entrants sur la majorité de la période montre une possible stagnation ou un manque d'intérêt pour le secteur public, peut-être dû à divers facteurs économiques, politiques ou institutionnels.

Opportunité d'analyse : Le pic de 2010 mérite une attention particulière pour comprendre ce qui a motivé cet afflux et s'il est possible de reproduire de telles conditions pour stimuler de futurs investissements.

1. **Importation des bibliothèques :**



Importation de matplotlib.pyplot : Cela permet de créer des visualisations graphiques. pyplot est souvent utilisé pour tracer des graphiques et des diagrammes.

Cette fonction est utilisée pour diviser un jeu de données en ensembles d'entraînement et de test, ce qui est essentiel pour évaluer la performance d'un modèle.

Importation de la régression linéaire : Cela permet de créer un modèle de régression linéaire, qui est utilisé pour prédire une variable continue à partir d'une ou plusieurs variables indépendantes.

Importation de métriques pour évaluer le modèle : mean\_squared\_error calcule l'erreur quadratique moyenne, mesurant la différence entre les valeurs prédites par le modèle et les valeurs réelles.

r2\_score mesure la proportion de la variance des résultats qui est prévisible à partir des variables indépendantes.

1. **Traitement de la régression linéaire :**



Sélection des colonnes pertinentes : Ici, un nouveau DataFrame data\_model est créé en sélectionnant uniquement les colonnes year, quarter, et Public\_PD\_inflows à partir de l'objet data. Cela permet de se concentrer sur les données pertinentes pour l'analyse.

Suppression des lignes avec des valeurs manquantes : Cette ligne supprime toutes les lignes contenant des valeurs manquantes dans data\_model, garantissant que le modèle soit entraîné uniquement sur des données complètes.

Définir les variables indépendantes et dépendantes : X contient les variables indépendantes (prédicteurs), ici les colonnes year et quarter. Y contient la variable dépendante (cible), qui est Public\_PD\_inflows.

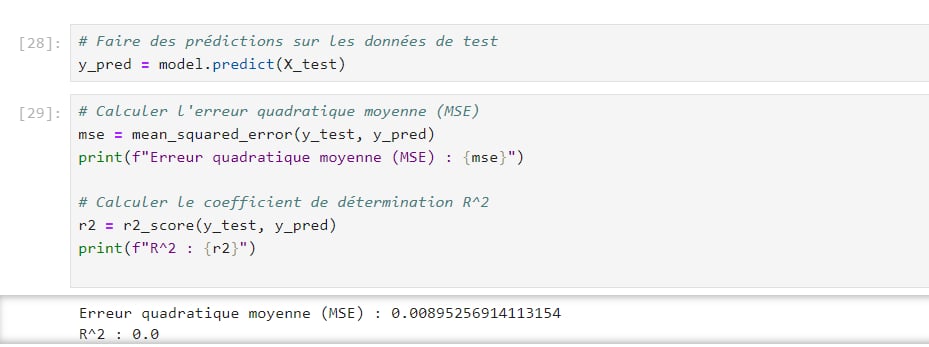
Diviser les données en ensembles d'entraînement et de test : Les données sont divisées en ensembles d'entraînement et de test. 80% des données sont utilisées pour entraîner le modèle (X\_train et y\_train), et 20% pour tester sa performance (X\_test et y\_test).

random\_state=42 fixe la graine pour garantir que les résultats soient reproductibles.

Créer le modèle de régression linéaire : Un objet model de type LinearRegression est créé, prêt à être entraîné sur les données.

Entraîner le modèle sur les données d'entraînement : Le modèle est entraîné en ajustant ses paramètres aux données d'entraînement (X\_train et y\_train).

1. **Calcul de la Prédiction :**



Faire des prédictions sur les données de test : Cette ligne utilise le modèle de régression linéaire entraîné pour faire des prédictions sur les données de test (X\_test). Les valeurs prédites sont stockées dans la variable y\_pred.

Calculer l'erreur quadratique moyenne (MSE) : Ici, l'erreur quadratique moyenne (MSE) est calculée en comparant les valeurs réelles (y\_test) avec les valeurs prédites (y\_pred). Le MSE mesure la moyenne des carrés des erreurs, ce qui indique la qualité des prédictions.

La valeur de MSE est ensuite affichée.

Calculer le coefficient de détermination R2 : Le coefficient de détermination 𝑅2 est calculé pour évaluer la proportion de la variance des résultats qui est expliquée par le modèle. Une valeur de 𝑅2 proche de 1 indique un bon ajustement du modèle, tandis qu'une valeur proche de 0 indique que le modèle ne parvient pas à expliquer la variance des données. La valeur de 𝑅2 est ensuite affichée.

Ce code effectue des prédictions sur les données de test, puis évalue la performance du modèle en calculant l'erreur quadratique moyenne (MSE) et le coefficient de détermination 𝑅2. Ces métriques aident à comprendre à quel point le modèle est performant dans ses prédictions.

1. Valeurs réelle vu prédites :



Configurer la taille de la figure : Cette ligne crée une nouvelle figure pour le graphique avec une taille de 10 unités de large et 6 unités de haut.

Tracer les valeurs réelles : Un nuage de points est créé pour représenter les valeurs réelles (y\_test) par rapport aux valeurs prédites (y\_pred).

Les points sont colorés en bleu avec une transparence de 0.6, ce qui permet une meilleure visualisation. Le label "Prédictions" est également ajouté pour la légende.

Tracer la ligne de prédiction idéale : Une ligne rouge est tracée pour représenter la ligne de prédiction idéale, où les valeurs prédites seraient égales aux valeurs réelles. Cette ligne sert de référence pour évaluer la qualité des prédictions.

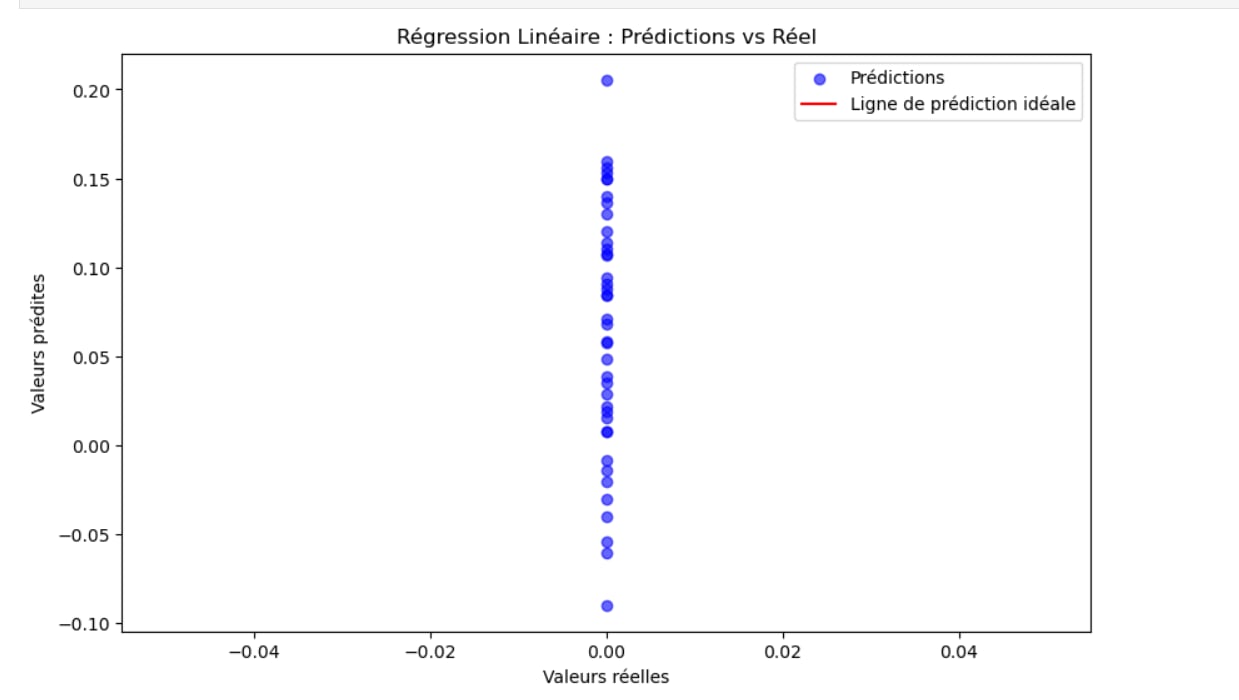
Titre et étiquettes : Le titre du graphique est défini comme "Régression Linéaire : Prédictions vs Réel".

Les axes sont étiquetés, avec "Valeurs réelles" pour l'axe des abscisses et "Valeurs prédites" pour l'axe des ordonnées.

Afficher la légende : Cette ligne affiche la légende du graphique, qui inclut les étiquettes définies précédemment pour les points et la ligne.

Afficher le graphique : enfin, cette ligne affiche le graphique à l'écran.

1. **Représentation de la régression linéaire :**



* Description du Graphique :

Titre :

Le graphique est intitulé "Régression Linéaire : Prédictions vs Réel", ce qui indique qu'il compare les valeurs prédites par le modèle de régression linéaire avec les valeurs réelles observées.

Axes :

Axe des abscisses (Valeurs réelles) : Représente les valeurs réelles des données de test.

Axe des ordonnées (Valeurs prédites) : Représente les valeurs prédictes par le modèle.

Points bleus :

Les points bleus représentent les valeurs prédites pour chaque valeur réelle. Dans ce graphique, tous les points semblent alignés verticalement, ce qui indique qu'il y a très peu de variation dans les valeurs prédites, indépendamment des valeurs réelles.

Ligne rouge :

La ligne rouge représente la "Ligne de prédiction idéale", où les valeurs prédites seraient égales aux valeurs réelles. Idéalement, tous les points devraient se situer sur cette ligne pour indiquer que le modèle fait des prédictions parfaites.

Légende :

La légende indique que les points bleus correspondent aux "Prédictions" et que la ligne rouge est la "Ligne de prédiction idéale".

Interprétation

* Proximité des points à la ligne idéale :

Les points ne sont pas proches de la ligne rouge, ce qui suggère que les prédictions du modèle ne sont pas satisfaisantes. Cela pourrait signifier que le modèle ne capture pas bien la relation entre les variables, ou qu'il y a un problème avec les données.

* Variabilité des valeurs prédites :

Le fait que les points soient presque tous alignés verticalement sur un intervalle très étroit indique une faible variabilité dans les prédictions. Cela pourrait être un signe de sur ajustement, où le modèle a appris trop de détails spécifiques des données d'entraînement, mais échoue à généraliser.

* Analyse des performances :

Ce graphique met en évidence une performance médiocre du modèle de régression linéaire. Une meilleure représentation des prédictions serait celle où les points seraient dispersés autour de la ligne rouge, indiquant que le modèle fait des prédictions raisonnablement proches des valeurs réelles.

**Conclusion**

Ce projet sur l'analyse des flux de capitaux internationaux en Tunisie a permis d'identifier des tendances et des problématiques clés dans le domaine économique et financier. À travers l'utilisation de techniques statistiques et de visualisation de données, nous avons pu explorer les inflows et outflows de capitaux, en mettant en lumière les valeurs manquantes et les anomalies présentes dans les données.

Synthèse des résultats

Analyse des Données : La majorité des flux d'inflows et d'outflows étaient faibles ou absents, ce qui soulève des questions sur l'activité économique et l'attractivité des investissements en Tunisie. Les valeurs nulles ou manquantes indiquent un manque de données fiables, nécessitant un nettoyage approfondi.

Implications Économiques : Les résultats montrent que la Tunisie est majoritairement un importateur net de capitaux, ce qui peut refléter une dépendance économique à l'égard des investissements étrangers. Les secteurs avec des inflows significatifs pourraient représenter des opportunités de croissance, tandis que ceux avec des outflows négatifs nécessitent une attention particulière.

Modélisation et Prédictions : L'application de modèles de régression linéaire a révélé des limites dans la capacité de prédiction, soulignant la nécessité d'améliorer la qualité des données et d'explorer des modèles plus complexes pour capturer les dynamiques des flux de capitaux.