
Rapport du Mini projet : Python/Pandas

éalisé par :

➤ SOULAIMANI Meryam

Année Universitaire : 2024 -2025

Introduction

Ce projet a pour objectif d'explorer et d'analyser un jeu de données d'étudiants en utilisant la bibliothèque Pandas en Python. Nous avons créé un programme qui importe, manipule, analyse et visualise les données d'étudiants, notamment leurs noms, prénoms, notes, filières et années de baccalauréat. Ce rapport présente les étapes du projet, les résultats obtenus et les visualisations réalisées.

Cahier des Charges

1. Données

- Source des données: Fichier CSV nommé `notes.csv`.
- Format attendu : Tableau de données avec les colonnes suivantes :
 - `Nom`
 - `Prénom`
 - `Note`
 - `Filière`
 - `Année_Bac`

```
import pandas as pd

# Créer un dictionnaire avec les données des étudiants
data = {
    "Nom": ["Dupont", "Martin", "Bernard", "Durand", "Leroy", "Moreau", "Garnier", "Petit", "Roux", "Lemoine"],
    "Prénom": ["Jean", "Sophie", "Luc", "Claire", "Paul", "Emma", "Julien", "Chloé", "Thomas", "Alice"],
    "Note": [15.5, 12.0, 18.0, 14.5, 10.0, 16.0, 11.5, 19.0, 13.0, 17.5],
    "Filière": ["Sciences", "Lettres", "Sciences", "Économie", "Lettres", "Sciences", "Économie", "Lettres", "Sciences", "Écono"],
    "Année de Bac": [2022, 2021, 2022, 2023, 2021, 2022, 2023, 2022, 2021, 2023]
}

# Créer un DataFrame à partir du dictionnaire
df = pd.DataFrame(data)

# Exporter Le DataFrame vers un fichier CSV
df.to_csv('notes.csv', index=False)

# Afficher un message pour confirmer que Le fichier a été créé
print("Le fichier CSV a été créé avec succès.")

Le fichier CSV a été créé avec succès.
```

2. Fonctionnalités Implémentées

1. **Importation des données** : Chargement du fichier CSV dans un DataFrame Pandas.

2. **Exploration des données** :

- Affichage des premières lignes du DataFrame.
- Vérification des informations générales (types de données, nombre de lignes et colonnes).
- Vérification et gestion des valeurs manquantes.

```
import pandas as pd
import matplotlib.pyplot as plt

# 1. Importation des données
df = pd.read_csv('notes.csv')

# 2. Exploration des données
print("Premières lignes du DataFrame :")
print(df.head())

print("\nInformations générales :")
print(df.info())

print("\nValeurs manquantes :")
print(df.isnull().sum())

# Gérer les valeurs manquantes (exemple : supprimer les lignes avec des valeurs manquantes)
df.dropna(inplace=True)
```

Premières lignes du DataFrame :

	Nom	Prénom	Note	Filière	Année de Bac
0	Dupont	Jean	15.5	Sciences	2022
1	Martin	Sophie	12.0	Lettres	2021
2	Bernard	Luc	18.0	Sciences	2022
3	Durand	Claire	14.5	Économie	2023
4	Leroy	Paul	10.0	Lettres	2021

Informations générales :

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 10 entries, 0 to 9
```

```
Data columns (total 5 columns):
```

#	Column	Non-Null Count	Dtype
0	Nom	10 non-null	object
1	Prénom	10 non-null	object
2	Note	10 non-null	float64
3	Filière	10 non-null	object

3. ****Manipulation des données**** :

- Filtrage des étudiants ayant une note supérieure à 15.
- Tri des données par note.

```
# 3. Manipulation des données
# Filtrer les étudiants ayant une note supérieure à 15
etudiants_bons = df[df['Note'] > 15]
print("\nÉtudiants ayant une note supérieure à 15 :")
print(etudiants_bons)

# Trier les données par note
df.sort_values(by='Note', ascending=False, inplace=True)

# 4. Analyse statistique
print("\nStatistiques descriptives :")
print(df.describe())
```

```
4  Année de Bac  10 non-null    int64
dtypes: float64(1), int64(1), object(3)
memory usage: 532.0+ bytes
None
```

Valeurs manquantes :

```
Nom          0
Prénom       0
Note         0
Filière      0
Année de Bac 0
dtype: int64
```

Étudiants ayant une note supérieure à 15 :

	Nom	Prénom	Note	Filière	Année de Bac
0	Dupont	Jean	15.5	Sciences	2022
2	Bernard	Luc	18.0	Sciences	2022
5	Moreau	Emma	16.0	Sciences	2022

4. ****Analyse statistique**** :

- Calcul des statistiques descriptives (moyenne, écart-type, etc.).
- Calcul de la moyenne des notes par filière.

```
# Calculer la moyenne des notes par filière
moyenne_notes_par_filiere = df.groupby('Filière')['Note'].mean()
print("\nMoyenne des notes par filière :")
print(moyenne_notes_par_filiere)
```

7	Petit	Chloé	19.0	Lettres	2022
9	Lemoine	Alice	17.5	Économie	2023

Statistiques descriptives :

	Note	Année de Bac
count	10.000000	10.000000
mean	14.700000	2022.000000
std	3.020302	0.816497
min	10.000000	2021.000000
25%	12.250000	2021.250000
50%	15.000000	2022.000000
75%	17.125000	2022.750000
max	19.000000	2023.000000

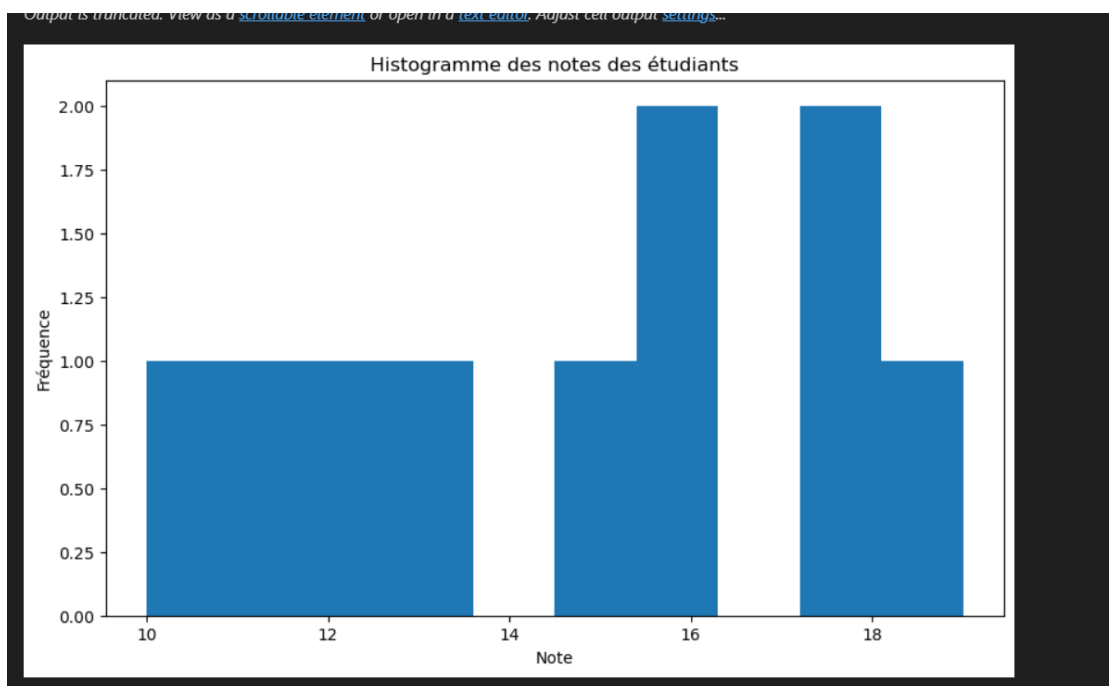
Moyenne des notes par filière :

Filière	
Lettres	13.666667
Sciences	15.625000

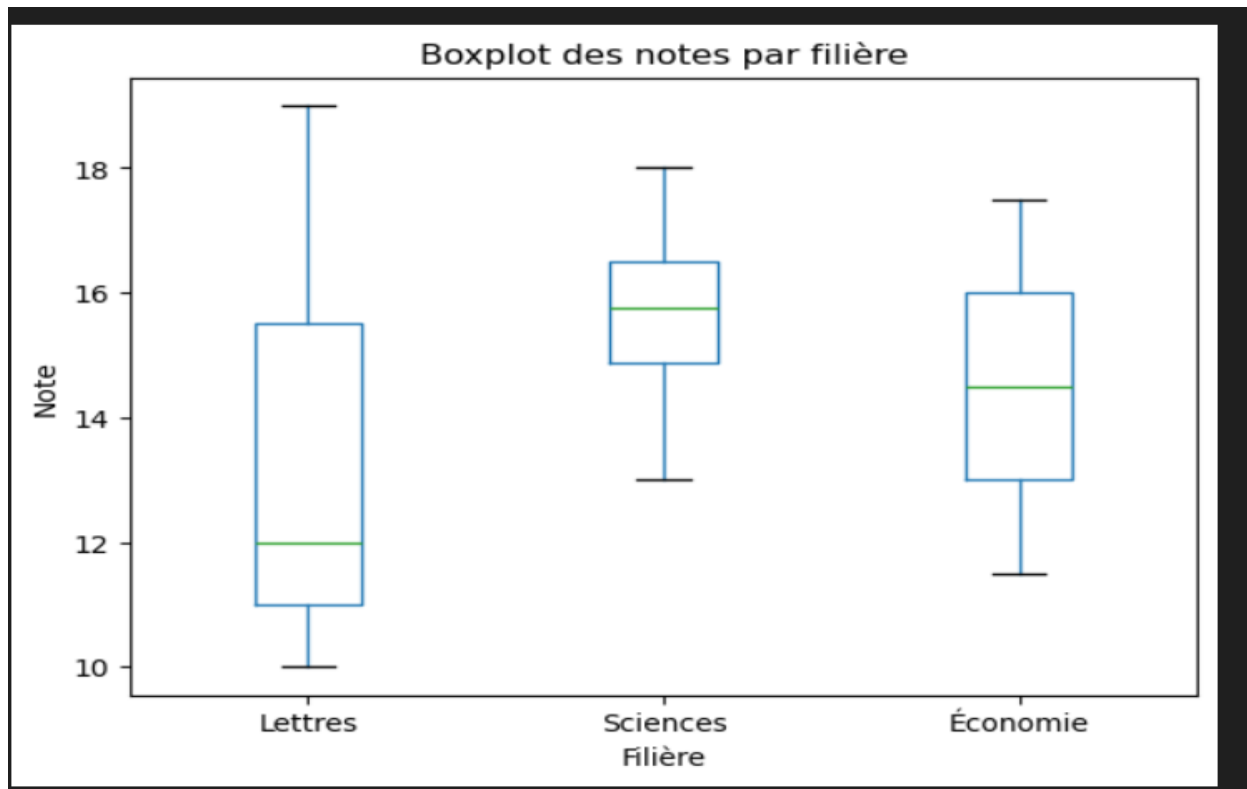
5. **Visualisation des données** :

- Histogramme des notes.

```
# 5. Visualisation des données
# Histogramme des notes
plt.figure(figsize=(10, 6))
df['Note'].hist(bins=10)
plt.title('Histogramme des notes des étudiants')
plt.xlabel('Note')
plt.ylabel('Fréquence')
plt.grid(False)
plt.show()
```

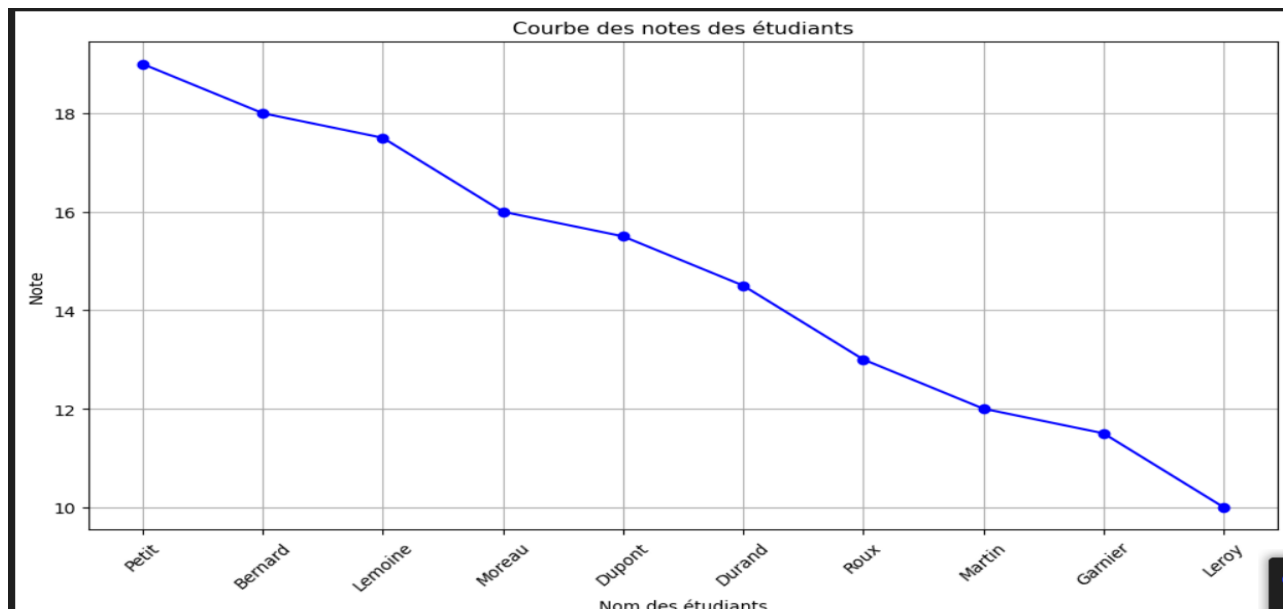


Boxplot des notes par filière.



- Courbe des notes des étudiants.

```
# Visualisation des courbes
plt.figure(figsize=(10, 6))
plt.plot(df['Nom'], df['Note'], marker='o', linestyle='-', color='b')
plt.title('Courbe des notes des étudiants')
plt.xlabel('Nom des étudiants')
plt.ylabel('Note')
plt.xticks(rotation=45) # Rotation des noms pour une meilleure lisibilité
plt.grid()
plt.tight_layout() # Ajuste le layout pour éviter le chevauchement
plt.show()
```



Conclusion

En somme, ce mini-projet a été une expérience enrichissante qui a permis de mettre en pratique des compétences en manipulation et analyse de données avec Pandas, ainsi qu'en visualisation avec Matplotlib. Les résultats obtenus ouvrent la voie à des analyses futures et à une meilleure compréhension des performances académiques des étudiants. Ce projet démontre également l'importance de l'analyse de données dans le domaine éducatif, où des décisions éclairées peuvent être prises pour améliorer les résultats des étudiants.

<https://github.com/MerySoulaïmani/Python-Pandas.git>