



Université Sultan Moulay Slimane Faculté
Polydisciplinaire – Beni Mellal



Filière : LICENCE D'EXCELLENCE

Spécialité : I3D et IA

Module : Python

Rapport de Mini Projet Pandas / Python

Réalisé par :

NOM : Hassan NACHIT

Numéro Apogée: 1307518

Encadré par :

Pr. MOUNCIF

Année universitaire : 2024-2025

Sommaire :

1. Introduction

2. Cahier des Charges

- 2.1. Données
- 2.2. Fonctionnalités à Implémenter

3. Étapes Réalisées

- 3.1. Importation des Données
- 3.2. Exploration des Données
- 3.3. Manipulation des Données
- 3.4. Analyse Statistique
- 3.5. Visualisation des Données

4. Conclusion

5. Liens et Références

1. Introduction :

Pandas est une bibliothèque Python très populaire pour la manipulation et l'analyse de données. Ce mini-projet a pour but de démontrer l'utilisation de Pandas à travers un jeu de données concernant les notes des étudiants. Nous allons importer, manipuler, analyser, et visualiser ces données

2. Cahier des Charges

2.1. Données

- **Source des données :** Fichier CSV.
- **Format attendu :** Tableau de données avec des colonnes pour les identifiants des étudiants, leurs noms, prénoms, matières, notes, classes, et années.

2.2. Fonctionnalités à Implémenter

1. Importation des données

- Charger un fichier CSV dans un DataFrame Pandas.

2. Exploration des données

- Afficher les premières lignes.
- Vérifier les informations générales et les valeurs manquantes.

3. Manipulation des données

- Sélectionner des colonnes, filtrer des lignes, créer de nouvelles colonnes, et trier les données.

4. Analyse statistique

- Calculer des statistiques descriptives et grouper les données.

5. Visualisation des données

- Tracer des graphiques simples à l'aide de Pandas et Matplotlib.

3. Étapes Réalisées :

3-1. Importation des Données :

Nous avons commencé par importer la bibliothèque Pandas et charger le fichier CSV dans un DataFrame.

Code python :

```
import pandas as pd
import matplotlib.pyplot as plt
try:
    df = pd.read_csv("notes_etudiants_100_lignes.csv")
    print("✅ Fichier chargé avec succès !")
except FileNotFoundError:
    print("❌ Erreur : Fichier introuvable. Vérifiez le chemin ou le nom.")
```

❌ Erreur : Fichier introuvable. Vérifiez le chemin ou le nom.

3-2. Exploration des Données :

Nous avons exploré les données pour comprendre leur structure et vérifier les valeurs manquantes.

Code python :

```
print("\n=== Premières lignes ===")
print(df.head())

print("\n=== Informations générales ===")
print(df.info())

print("\n=== Valeurs manquantes ===")
print(df.isnull().sum())
```

Résultats :

```
=== Premières lignes ===
   id_etudiant  nom  prenom  matiere  note  classe  annee
0            1  Dupont   Jean  Mathématiques    15  Terminale  2023
1            2  Martin  Sophie    Physique    12  Terminale  2023
2            3  Bernard  Pierre  Mathématiques     8   Première  2023
3            4   Petit   Lucie    Physique    17  Terminale  2023
4            5  Moreau   Thomas  Mathématiques    10   Première  2023
```

```
=== Informations générales ===
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   id_etudiant     100 non-null   int64
1   nom             100 non-null   object
2   prenom          100 non-null   object
3   matiere         100 non-null   object
4   note            100 non-null   int64
5   classe          100 non-null   object
6   annee           100 non-null   int64
dtypes: int64(3), object(4)
memory usage: 5.6+ KB
```

```
=== Valeurs manquantes ===
id_etudiant    0
nom            0
prenom         0
matiere        0
note           0
classe         0
annee          0
dtype: int64
```

3.3. Manipulation des Données :

Nous avons sélectionné des colonnes spécifiques et créé une nouvelle colonne pour la moyenne des notes.

Code python :

```
df['mention'] = df['note'].apply(
    lambda x: 'Excellent' if x >= 16 else
              'Très bien' if x >= 14 else
              'Bien' if x >= 12 else
              'Passable' if x >= 10 else
              'Insuffisant'
)
# Filtrage des étudiants de Terminale
terminale = df[df['classe'] == 'Terminale']
```

3.4. Analyse Statistique :

Nous avons calculé des statistiques descriptives et groupé les données par classe.

Code python :

```
print("\n=== Statistiques par matière ===")
print(df.groupby('matiere')['note'].describe())

print("\n=== Moyenne par classe ===")
print(df.groupby('classe')['note'].mean().round(2))
```

Résultats :

```
=== Statistiques par matière ===
          count  mean      std  min   25%   50%   75%   max
matiere
Mathématiques   50.0  12.4  3.037185  8.0  10.0  13.0  15.0  16.0
Physique        50.0  12.6  2.755329  9.0  11.0  12.0  14.0  17.0

=== Moyenne par classe ===
classe
Première      9.0
Terminale     14.0
Name: note, dtype: float64
```

3.5. Visualisation des Données :

Nous avons utilisé Matplotlib pour tracer un histogramme des notes.

Code python :

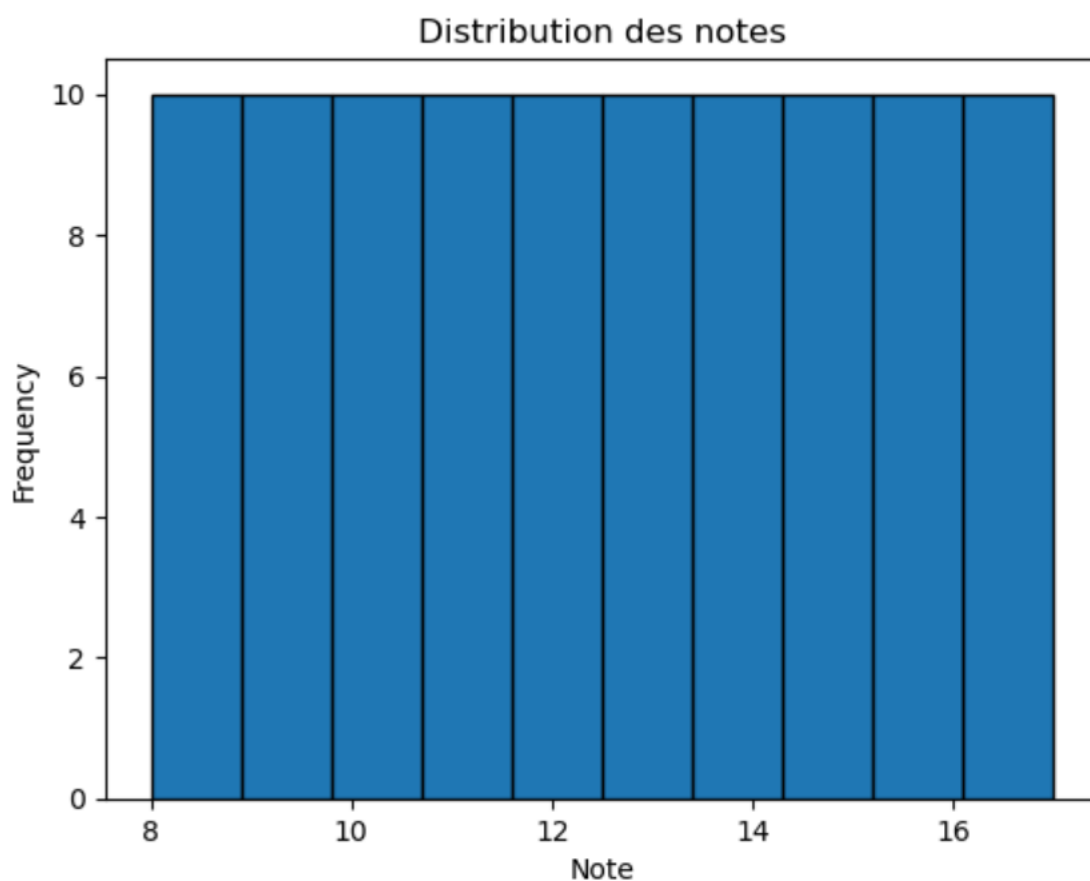
```

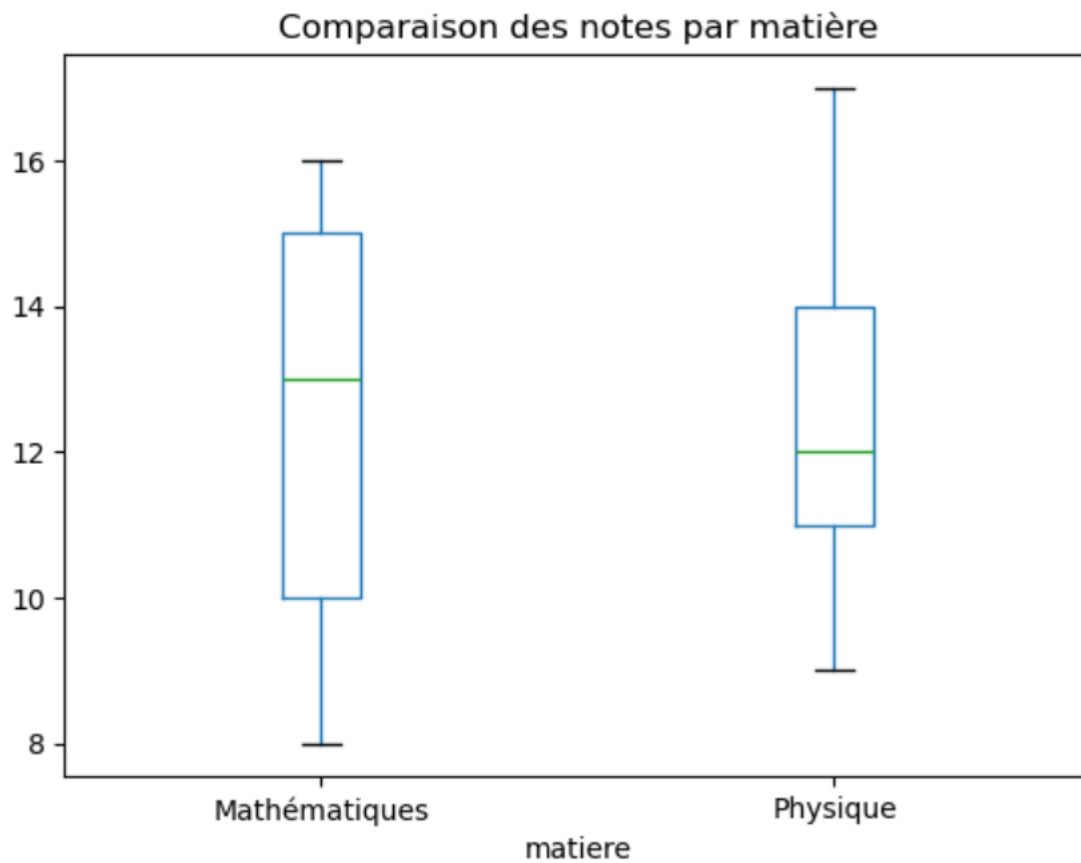
df['note'].plot(
    kind='hist',
    bins=10,
    title='Distribution des notes',
    edgecolor='black'
)
plt.xlabel('Note')
plt.show()

df.boxplot(column='note', by='matiere', grid=False)
plt.title('Comparaison des notes par matiere')
plt.suptitle('') # Supprime le titre automatique
plt.show()

```

Résultats :





3.5. Export des résultats :

```
df.to_csv('resultats_analyse.csv', index=False)
print("\n Résultats exportés dans 'resultats_analyse.csv'")
```

Résultats exportés dans 'resultats_analyse.csv'

4. Conclusion

Ce mini-projet a permis de mettre en pratique les compétences en manipulation et analyse de données avec Pandas. Nous avons importé des données, effectué des analyses descriptives, et visualisé les résultats de manière efficace. Les compétences acquises seront utiles pour des projets futurs en science des données.