Atelier : Premiers pas avec Pandas

Objectif
Cet atelier vise à vous familiariser avec la bibliothèque Pandas, un outil essentiel pour la manipulation et l'analyse de données en Python, particulièrement dans le contexte du machine learning.

Prérequis
- Python installé (version 3.7 ou supérieure)
- Pandas installé (`pip install pandas`)
- Jupyter Notebook ou un environnement de développement Python

Exercice 1 : Création et Manipulation de DataFrames

Vous allez créer un DataFrame représentant les informations d'employés d'une entreprise. Votre mission est de :

- Créer un DataFrame à partir d'un dictionnaire de données
- Explorer les premières lignes du dataset
- Calculer des statistiques de base
- Filtrer des données selon des conditions spécifiques

Exemple:

```python
import pandas as pd

# Création d'un DataFrame à partir d'un dictionnaire
data = {
    'Nom': ['Alice', 'Bob', 'Charlie', 'David'],
    'Âge': [25, 30, 35, 28],
    'Salaire': [50000, 60000, 75000, 55000],
    'Département': ['RH', 'Tech', 'Finance', 'Marketing']
}
df = pd.DataFrame(data)
```

Questions:
1. Affichez les 3 premières lignes du DataFrame
2. Calculez l'âge moyen des employés
3. Filtrez les employés qui gagnent plus de 55000

Exercice 2 : Analyse Statistique Descriptive

Approfondissez votre analyse statistique en utilisant les méthodes intégrées de Pandas. Vous allez :

- Générer un résumé statistique complet du DataFrame
- Regrouper les données par catégorie
- Calculer des agrégations

Exemple d'utilisation:

1. Calculez les statistiques descriptives du DataFrame
print(df.describe())
2. Groupez les données par département et calculez le salaire moyen
salaire_par_departement = df.groupby('Département')['Salaire'].mean()
print(salaire_par_departement)

Exercice 3 : Manipulation et Transformation de Données

Travaillez sur un jeu de données de ventes. Vous allez, premièrement, importer des données depuis un fichier CSV. Par la suite, vous nettoyez et préparez les données et vous créez de nouvelles colonnes calculées et enfin vous triez les données.

Exemple:

```
# Importation de données depuis un fichier CSV
#Supposons que vous ayez un fichier 'donnees_ventes.csv'
df_ventes = pd.read_csv('donnees_ventes.csv')
```

Tâches:

```
# 1. Gérez les valeurs manquantes
df_ventes.dropna(inplace=True)

# 2. Créez une nouvelle colonne calculée
df_ventes['Marge'] = df_ventes['Prix_Vente'] - df_ventes['Coût_Production']

# 3. Triez le DataFrame par marge décroissante
df_ventes_tries = df_ventes.sort_values('Marge', ascending=False)
```

Exercice 4 : Visualisation des Données

Utilisez Pandas et Matplotlib pour visualiser les données. Vous commencez par créer un histogramme des salaires, puis générer un graphique à barres des salaires par département et interpréter ces visualisations.

Exemple:

```
import matplotlib.pyplot as plt
# Visualisez la distribution des salaires
plt.figure(figsize=(10, 6))
df['Salaire'].hist(bins=10)
plt.title('Distribution des Salaires')
plt.xlabel('Salaire')
plt.ylabel('Fréquence')
plt.show()
# Créez un graphique à barres des salaires moyens par département
df.groupby('Département')['Salaire'].mean().plot(kind='bar')
plt.title('Salaire Moyen par Département')
plt.show()
```

Exercice 5 : Filtrage et Sélection Avancés

Maîtrisez les techniques de filtrage complexe et de sélection de colonnes. Vous allez :

- Filtrer des données avec plusieurs conditions
- Sélectionner des sous-ensembles spécifiques
- Manipuler des colonnes

Exemple:

```
# Filtres complexes
employes_selectionnes = df[
    (df['Âge'] > 25) &
    (df['Salaire'] > 55000) &
    (df['Département'] != 'RH')
]
# Sélection de colonnes spécifiques
infos_essentielles = df[['Nom', 'Département', 'Salaire']]
```