



Université Sultan Moulay Slimane Faculté Polydisciplinaire Béni Mellal
Département INFORMATIQUE (MIP)

Filière : Science de données et sécurité des systèmes
d'information
A.U : 2023-2024

Sujet

Compte Rendu de TP04 – Python pour la science des données

Présenté Par :
MAFTOUH Omar

Encadré Par :
Z. BOUSALEM
Y. MADANI

Introduction :

Dans ce travail pratique, nous explorerons deux exercices utilisant Python et la bibliothèque NumPy pour effectuer des analyses sur des données de ventes mensuelles de produits. Le premier exercice implique la création d'un tableau de données représentant les ventes de trois produits sur une année, suivi de divers calculs statistiques tels que le total des ventes, la moyenne mensuelle, l'identification des mois de ventes maximales et la croissance mensuelle en pourcentage. Le deuxième exercice se concentre sur la génération de données aléatoires pour deux produits, suivie d'analyses telles que le calcul de l'écart-type et du coefficient de corrélation entre les ventes des deux produits.

Exercice01 :

Dans cet exercice, nous explorerons un ensemble de données représentant les ventes mensuelles de trois produits (P1, P2 et P3) sur une période d'un an. Les tâches incluront la création d'un tableau 2D à l'aide de la bibliothèque NumPy, la génération de ventes aléatoires pour chaque mois et chaque produit, ainsi que l'application de diverses analyses statistiques. Nous calculerons le total des ventes pour chaque produit sur toute l'année, la moyenne des ventes mensuelles, et identifierons les mois de ventes maximales pour chaque produit. En outre, nous évaluerons la croissance mensuelle en pourcentage par rapport au mois précédent et déterminerons le mois présentant la plus forte croissance pour chaque produit. Enfin, nous créerons un tableau contenant la somme des ventes pour chaque mois, agrégeant ainsi les ventes de tous les produits.

```
import numpy as np

# Creation d'un tab Numpy 2d qui contener 12 ligne 3 colone , avce des valeur
aléatoire entre 100 & 1000
tab2D = np.random.randint(100 , 1000 , size=(12 , 3))
print(f"Tab 2D : \n{tab2D}")

print('*' * 50)

# Calculer et afficher le total des ventes pour chaque produit sur l'ensemble
de l'année
print(f"Total des ventes du P1 : {np.sum(tab2D[ : , 0])}")
print(f"Total des ventes du P2 : {np.sum(tab2D[ : , 1])}")
print(f"Total des ventes du P3 : {np.sum(tab2D[ : , 2])}")

print('*' * 50)

# Calculer et afficher la moyenne des ventes mensuelles pour chaque produit
print(f"Moyenne des ventes du P1 : {np.round(np.mean(tab2D[ : , 0]) , 2)}")
print(f"Moyenne des ventes du P2 : {np.round(np.mean(tab2D[ : , 1]) , 2)}")
print(f"Moyenne des ventes du P3 : {np.round(np.mean(tab2D[ : , 2]) , 2)}")

print('*' * 50)

# L'indentification le mois ayant les ventes maximales pour chaque produit
print(f"Le mois ayant les ventes maximales P1 : {np.argmax(tab2D[ : , 0]) +
1}")
print(f"Le mois ayant les ventes maximales P2 : {np.argmax(tab2D[ : , 1]) +
1}")
```

```

print(f"Le mois ayant les ventes maximales P3 : {np.argmax(tab2D[ : , 2]) + 1}")

print('*' * 50)

# Calculer la croissance mensuelle
TauxCroissance = ( np.diff(tab2D , axis=0) / tab2D[:-1] ) * 100
print(f"Taux de croissance du 3 Produits : \n{ np.int64(TauxCroissance) }")

print('*' * 50)

# Identifier et afficher le mois avec la plus forte croissance pour chaque produit
print(f"le mois le plus forte croissance pour P1 : { np.argmax(TauxCroissance[ : , 0])+1 }")
print(f"le mois le plus forte croissance pour P2 : { np.argmax(TauxCroissance[ : , 1])+1 }")
print(f"le mois le plus forte croissance pour P3 : { np.argmax(TauxCroissance[ : , 2])+1 }")

print('*' * 50)

# La creation d'un tableau 1D contenant la somme des ventes pour chaque mois
tab1D = np.array(np.sum(tab2D , axis=1))
print(f"Somme des ventes pour chaque mois : \n{tab1D}")

```

Résultat du programme :

```
Tab 2D :
[[171 474 900]
 [434 667 515]
 [689 815 488]
 [456 606 220]
 [237 351 183]
 [745 258 483]
 [538 824 962]
 [877 732 764]
 [227 397 611]
 [665 788 464]
 [547 406 752]
 [250 943 620]]
*****
Total des ventes du P1 : 5836
Total des ventes du P2 : 7261
Total des ventes du P3 : 6962
*****
Moyenne des ventes du P1 : 486.33
Moyenne des ventes du P2 : 605.08
Moyenne des ventes du P3 : 580.17
*****
Le mois ayant les ventes maximales P1 : 8
Le mois ayant les ventes maximales P2 : 12
Le mois ayant les ventes maximales P3 : 7
*****
Taux de coissance du 3 Produits :
[[153  40 -42]
 [ 58  22  -5]
 [-33 -25 -54]
 [-48 -42 -16]
 [214 -26 163]
 [-27 219  99]
 [ 63 -11 -20]
 [-74 -45 -20]
 [192  98 -24]
 [-17 -48  62]
 [-54 132 -17]]
*****
le mois le plus forte croissance pour P1 : 5
le mois le plus forte croissance pour P2 : 6
le mois le plus forte croissance pour P3 : 5
*****
Somme des ventes pour chaque mois :
[1545 1616 1992 1282  771 1486 2324 2373 1235 1917 1705 1813]
```

Exercice 02 :

Cet exercice se concentre sur l'analyse de la relation entre deux variables, X et Y, représentant les ventes mensuelles de deux produits distincts. Nous utiliserons des données générées de manière aléatoire pour chaque produit, en supposant des distributions normales pour les ventes. Nous explorerons ensuite la notion d'écart-type et son interprétation dans le contexte des ventes mensuelles. Nous calculerons les écarts-types pour X et Y, puis examinerons comment ces valeurs indiquent la variabilité des ventes pour chaque produit. En outre, nous étudierons la corrélation entre X et Y, en expliquant comment elle est calculée et comment interpréter son résultat pour comprendre la relation entre les ventes des deux produits. Enfin, nous examinerons comment ces mesures statistiques aident à saisir la dynamique des ventes de produits X et Y sur une période donnée.

```
import numpy as np

# Génération des données aléatoires pour X et Y
X = np.random.normal(loc=100 , scale=20, size=(1 , 12))
Y = np.random.normal(loc=120 , scale=25, size=(1 , 12))

# Calcul de l'écart-type pour X et Y
ecartTypeX = np.std(X)
ecartTypeY = np.std(Y)
print(f"Ecart Type du X : {ecartTypeX}")
print(f"Ecart Type du Y : {ecartTypeY}")

"""
    Pour X, un écart-type de 20 signifie que les ventes mensuelles varient
d'environ 20 unités autour de la moyenne.
    Pour Y, un écart-type de 25 signifie que les ventes mensuelles varient
d'environ 25 unités autour de la moyenne.
"""

# Calcul du coefficient de corrélation entre X et Y
CoeCorrelationX = np.corrcoef(X , Y)
print(f"le coefficient de correlation : \n{CoeCorrelationX}")
# -1 < Correlation < 1

"""
    * Un coefficient de corrélation proche de 1 indiquerait une corrélation
positive, c'est-à-dire que lorsque les ventes mensuelles de X augmentent, les
ventes mensuelles de Y ont tendance à augmenter également.
    * Un coefficient de corrélation proche de -1 indiquerait une corrélation
négative, ce qui signifie que lorsque les ventes mensuelles de X augmentent,
les ventes mensuelles de Y ont tendance à diminuer.
    * Un coefficient de corrélation proche de 0 indiquerait une faible
corrélation linéaire entre les ventes mensuelles de X et de Y.
"""

# 4

"""
    Un écart type plus élevé indique une plus grande variabilité des ventes
mensuelles.
```

```
Ainsi, un écart-type élevé pour un produit indique que les ventes
fluctuent davantage autour de la moyenne,
ce qui peut indiquer une demande plus volatile ou des facteurs externes
affectant les ventes de manière significative.
"""

# 5

"""
Le coefficient de corrélation mesure la relation linéaire entre les ventes
de X et Y.
Un coefficient positif indique une tendance à augmenter ensemble, tandis
qu'un coefficient négatif indique une tendance inverse.
Un coefficient proche de zéro indique une faible relation linéaire.
En utilisant ce coefficient, on peut évaluer l'ampleur et la signification
de la relation entre les ventes de X et Y.
"""
```

Résultat du programme :

```
Ecart Type du X : 17.25078030964672
Ecart Type du Y : 31.680566505260423
le coefficient de corrélation :
[[ 1.          -0.35638861]
 [-0.35638861  1.          ]]
```

Conclusion :

Ce travail pratique nous a permis de manipuler des données de ventes mensuelles de produits à travers deux exercices distincts. Nous avons utilisé des techniques statistiques telles que le calcul des totaux, des moyennes, des écarts-types et des coefficients de corrélation pour mieux comprendre les données et les relations entre les variables étudiées. Ces analyses fournissent des informations précieuses pour les décideurs dans le domaine du marketing et de la gestion des stocks.