

```

# -*- coding: utf-8 -*-
# Attention, sauf exception tout est écrit en anglais dans ce programme.

# print signifie imprimer.
# dataset signifie base de données.
# wage signifie salaire.
# CPI signifie Consumption Price Index, autrement dit, Indice des Prix à la Consommation.

# Attention, suivant les éditeurs de texte et les versions de windows,
# vous pouvez avoir des problèmes avec les accents dans les print.
# Il est donc conseillé d'écrire sans accent à l'intérieur des commandes print
# si vous avez un problème en exécutant ce programme.

# Utiliser plutôt print ('xxx') que print ("xxx"), sauf si vous avez une apostrophe dans le
# texte.

# Comme vous l'avez déjà constaté le # sert à mettre entre commentaires.
# Pour exécuter le programme, cliquer sur le triangle.

print ('Hello Anaconda')

import pickle

# Ouverture de la base de données wooldridge23
# Attention, il faudra remplacer le chemin C:/Users/claire.loupias/Desktop/
# dans la ligne ci-dessous par votre chemin à vous.
# rb signifie read binary.

fid=open('C:/Users/claire.loupias/Desktop/wooldridge23.pkl','rb')
alldatasets=pickle.load(fid)
fid.close()

#EXERCICE 1.1
print ('Exercice 1.1')

# Lecture de la base de données nécessaire pour l'exercice 1.1
dataset=alldatasets['wage1']

# Pour voir la structure de la base de données, taper dans la console les instructions
# suivantes :
# - dataset          Cela permet d'imprimer le début et la fin de la base de données.
# - list(dataset)   Cela permet d'avoir la liste des variables présentes dans la table.
# - dataset.shape   Cela permet de connaître le nombre d'observations et le nombre de variables.
# Attention, la première observation est numérotée zéro.

# Question 1
# le nom de la variable dans la base de données est educ.
education=dataset.educ          # Sélection de la variable.
moyenne=education.mean()       # Calcul de la moyenne.
print('Niveau moyen des études : '+f'{moyenne:.2f}'+ ' années')
mineduc=education.min()        # Calcul du minimum.
maxeduc=education.max()        # Calcul du maximum.
print('Minimum des années : '+str(mineduc))
print('Maximum des années : '+str(maxeduc))

# Question 2
# Le nom de la variable dans la base de données est wage.
salaire=dataset.wage
moyenne=salaire.mean()
print('Salaire moyen en 1976 : '+f'{moyenne:.2f}'+ ' dollars de 1976')

# Question 3
# Lecture des données macro dans la base IMF_CPI_A.US.PCPI_IX.csv
import pandas as pd
data=pd.read_csv('IMF_CPI_A.US.PCPI_IX.csv') # On récupère l'indice des prix à la consommation.
data.columns=['period','cpi']              # On renomme les colonnes (la seconde) afin de
                                           # faciliter la suite.

```

```

cpi1976=data[data.period==1976].cpi.values[0]
cpi2019=data[data.period==2019].cpi.values[0]

inflationFactor=cpi2019/cpi1976 #Facteur d'inflation
salaire1976_2019=moyenne*inflationFactor
print('Salaire moyen en 1976 : '+f'{salaire1976_2019:.2f}'+ ' en dollars de 2019')

# Question 4
# Le nom de la variable utilisée dans la bse de données est female.
female=dataset.female
n=sum(female) # Calcul du nombre de femmes dans l'échantillon.
m=female.size-n # Nombre d'hommes par complémentarité avec le nombre d'observations pris sur la
                # taille de la variable
print('Nombre de femmes : '+str(n))
print("Nombre d'hommes : "+str(m))
p=100*n/(n+m)
q=100-p
print('Proportion de femmes : '+f'{p:.2f}'+ ' %')
print('Proportion des hommes : '+f'{q:.2f}'+ ' %')

# Question 5
wmf=salaire[female==1].mean() # Salaire moyen des femmes.
wmh=salaire[female==0].mean() # Salaire moyen des hommes.
ratio=wmh/wmf
print('Le salaire horaire moyen des femmes en 1976 est de : '+f'{wmf:.2f}' + ' dollars')
print('Le salaire horaire moyen des hommes en 1976 est de : '+f'{wmh:.2f}'+ ' dollars')
print('Le salaire moyen des hommes est donc '+f'{ratio:.2f}'+ ' fois plus important.')

```