

```

1  '''
2  Correction de l'activité Capytale2 "Analyse de données CSV"
3  Auteur: Marc Biver
4  '''
5
6  def Sep():
7      '''
8      Fonction affichant une ligne à l'écran pour séparer
9      visuellement les réponses aux questions
10     Note: "\n" = passage à la ligne
11     '''
12     print("\n#####\n")
13
14     def PrtQ(num):
15         '''
16         Fonction qui affiche simplement le numéro de la question en cours,
17         qu'elle prend en entrée.
18         '''
19         print("Question(s):", num)
20         print("\n")
21
22     import csv # On va évidemment utiliser le module CSV
23
24     # Question 1 - ouverture en lecture du fichier
25     PrtQ(1)
26     fichier = open("DataLycees2022.csv","r",encoding = "utf-8")
27     print("Fichier ouvert")
28     Sep()
29
30     # Question 2 - chargement du contenu du fichier dans une
31     #               liste de dictionnaires
32     PrtQ(2)
33     table = list(csv.DictReader(fichier))
34     print("Fichier chargé")
35     Sep()
36
37     # Question 3.1 - affichage 5 premières lignes du fichier
38     PrtQ("3.1")
39     for i in range(5):
40         print(table[i])
41     print("5 premières lignes du fichier")
42     Sep()
43
44     #Question 3.2 - affichage des lignes 1500 à 1510
45     PrtQ("3.2")
46     for i in range(1499,1510):
47         print(table[i])
48     print("Lignes 1500 à 1510 du fichier")
49     Sep()
50
51     #Question 3.3 - nombre de lignes du fichier (hors ligne de titres de colonnes)
52     PrtQ("3.3")
53     print("Nombre lignes du fichier:", len(table))
54     Sep()
55
56     #Question 4.1 - département de commune d'Ambert
57     PrtQ("4.1")
58     for i in range(len(table)):
59         if table[i]['commune']=='AMBERT':
60             print("Département commune Ambert:", table[i]['departement'])

```

```

61 Sep()
62
63 #Question 4.2 - nombre de lycées en Nouvelle Aquitaine
64 PrtQ("4.2")
65 a = 0
66 for i in range(len(table)):
67     if table[i]['region_academique']=='NOUVELLE-AQUITAINE':
68         a+=1
69 print("Nombre lycées en Nouvelle-Aquitaine:", a)
70 Sep()
71
72 #Question 4.3 & 4.4 - recherche de noms de lycées
73 PrtQ("4.3 & 4.4")
74 def cherchelycee(patro, table):
75     '''
76     Fonction qui cherche dans la table dans la colonne patronyme le nom passé en
77     ↪ argument.
78     Si elle le trouve elle renvoie le nom de la commune.
79     Dans le cas contraire elle renvoie le message "Pas trouvé!")
80     '''
81     trouve = False
82     for i in range (len(table)):
83         if table[i]['patronyme'] == patro:
84             return(table[i]['commune'])
85             trouve = True
86
87     if trouve == False:
88         return('Pas trouvé!')
89 print("Commune lycée Docteur Dupont:", cherchelycee("DOCTEUR DUPONT", table))
90 print("Commune lycée Docteur Dupont:", cherchelycee("DOCTEUR KOEBERLE", table))
91 Sep()
92
93 #Questions 5.1, 5.2, 5.3 - élèves en 2nde GT au lycée général de Uturoa
94 PrtQ("5.1, 5.2, et 5.3")
95 for i in range(len(table)):
96     if table[i]['commune'] == 'UTUROA' and \
97         table[i]['denomination_principale'] == 'LYCEE GENERAL' :
98         Nb2nde = int(table[i]['2ndes_gt'])
99         NbFilles = int(table[i]['2ndes_gt_filles'])
100 print("Nombre élèves 2ndeGT Lycée Général Uturoa:", Nb2nde)
101 print("Nombre filles 2ndeGT Lycée Général Uturoa:", NbFilles)
102 PropFilles = round(NbFilles / Nb2nde * 100, 2)
103 print("Proportion filles 2ndeGT Lycée Général Uturoa:", PropFilles, "%")
104 Sep()
105
106 #Question 6.1 - Validation des données de la table
107 # Remarque: pour un problème d'encodage, le titre de la première colonne
108 # ("rentrée scolaire") ne passe pas tel quel - il faut utiliser
109 ↪ "\ufeffrentrée_scolaire"
110 # à la place.
111 # (pour vraiment résoudre le problème il aurait fallu utiliser l'encodage
112 # "utf-8-sig" au moment de la lecture du fichier - mais on ne va pas
113 # rentrer là-dedans ici.
114 PrtQ("6.1")
115 def validation_table(table_donnees):
116     '''
117     Fonction de validation des données de rentrées scolaires des lycées de
118     ↪ France.
119     Concrètement, conversion au format entier de toutes les données entières.
120     '''

```

```

118     for i in range(len(table_donnees)):
119         table_donnees[i]['\u00e9ffrentree_scolaire'] =
120             ↪ int(table_donnees[i]['\u00e9ffrentree_scolaire'])
121         table_donnees[i]['nombre_d_eleves'] =
122             ↪ int(table_donnees[i]['nombre_d_eleves'])
123         table_donnees[i]['2ndes_gt'] = int(table_donnees[i]['2ndes_gt'])
124         table_donnees[i]['2ndes_gt_filles'] =
125             ↪ int(table_donnees[i]['2ndes_gt_filles'])
126         table_donnees[i]['2ndes_gt_garcons'] =
127             ↪ int(table_donnees[i]['2ndes_gt_garcons'])
128         table_donnees[i]['2ndes_gt_lv1_allemand'] =
129             ↪ int(table_donnees[i]['2ndes_gt_lv1_allemand'])
130         table_donnees[i]['2ndes_gt_lv1_anglais'] =
131             ↪ int(table_donnees[i]['2ndes_gt_lv1_anglais'])
132         table_donnees[i]['2ndes_gt_lv1_espagnol'] =
133             ↪ int(table_donnees[i]['2ndes_gt_lv1_espagnol'])
134     return table_donnees
135
136 table = validation_table(table) # donc à partir de maintenant toutes les données
137 ↪ seront au bon format
138 print("Données de la table validées")
139 Sep()
140
141 #Question 6.2 - pourcentage de filles en 2nde GT
142 PrtQ("6.2")
143 def pourcent_fille(table, deno, patro):
144     '''
145     Fonction qui prend en entrée une table de données, la dénomination
146     ↪ principale
147     et le patronyme d'un lycée et qui renvoie le pourcentage de filles en 2nde
148     ↪ GT
149     '''
150     nb2nde = 0
151     nbFilles = 0
152     for i in range(len(table)):
153         if table[i]['denomination_principale'] == deno and table[i]['patronyme']
154             ↪ == patro:
155             nbFilles += table[i]['2ndes_gt_filles']
156             nb2nde = table[i]['2ndes_gt'] # pas besoin de int() parce que la
157             ↪ validation des données a été faite
158     return(round((nbFilles/nb2nde)*100,2))
159
160 # 2 tests de cette fonction:
161 print("Pourcentage de filles en 2nde au Lycée Général Lamartine", \
162       pourcent_fille(table, "LYCEE GENERAL", "LAMARTINE"), "%")
163 print("Pourcentage de filles en 2nde au Lycée Polyvalent Geneviève Vincent", \
164       pourcent_fille(table, 'LYCEE POLYVALENT', 'GENEVIEVE VINCENT'), "%")
165 Sep()
166
167 #Question 7.1 - nombre d'élèves en 2ndeGT en France à la rentrée 2022
168 PrtQ("7.1")
169 somme = 0
170 for i in range(len(table)):
171     if table[i]['\u00e9ffrentree_scolaire'] == 2022:
172         somme += table[i]['2ndes_gt']
173 print("Nombre total d'élèves en 2ndeGT en France à la rentrée 2022:", somme)
174 Sep()
175
176 #Question 7.2 - pourcentage du nombre total d'élèves en établissement privé et
177 ↪ ont choisi allemand comme LV1

```

```

165 PrtQ("7.2")
166 nb2nde = 0
167 nbPriveAllemand = 0
168 for i in range(len(table)):
169     nb2nde += table[i]['2ndes_gt']
170     if table[i]['secteur'] == 'PRIVE':
171         nbPriveAllemand += table[i]['2ndes_gt_lv1_allemand']
172
173 pct = round((nbPriveAllemand/nb2nde)*100,2)
174 print("% d'élèves en établissement privé et ont choisi allemand comme LV1:",
175       ↪ pct, "%")
176 Sep()
177 #Question 7.3 - pourcentage du nombre total d'élèves qui sont en établissement
178 ↪ privé, ont choisi allemand comme LV1, et sont des filles?
179 PrtQ("7.3")
180 print("Cette question était un piège - voir le commentaire dans le code du
181 ↪ corrigé")
182 # Les données ne permettent pas de répondre à cette question: par établissement
183 ↪ on connaît le nombre de filles,
184 # et on connaît le nombre d'élèves qui ont pris allemand comme LV1, mais on ne
185 ↪ peut pas connaître l'intersection
186 # des deux - combien, parmi les filles, ont fait ce choix de LV1.
187 # Il était donc impossible de répondre à la question.
188 Sep()
189
190 #Question 7.4 - pourcentage moyen de filles en 2nde GT en établissement privé?
191 ↪ En établissement public? Au total?
192 PrtQ("7.4")
193 # Cette question est un peu plus complexe puisqu'on parle de pourcentage moyen -
194 ↪ ce qui en soit est ambigu.
195 # On va donc calculer ici deux choses pour les comparer:
196 # Le pourcentage moyen (donc la somme de toutes les filles sur la somme de tous
197 ↪ les élèves) et la
198 # moyenne des pourcentages (chaque pourcentage de chaque établissement étant
199 ↪ pris dans une liste dont
200 # on va faire la moyenne).
201 # Pour rendre le code lisible on va utiliser un dictionnaire de valeurs
202 sommes = dict()
203 sommes['2nde-Public'] = 0 # Nombre d'élèves en 2nde dans le public
204 sommes['2nde-Privé'] = 0 # Nombre d'élèves en 2nde dans le privé
205 sommes['Filles-Public'] = 0 # Nombre d'élèves filles dans le public
206 sommes['Filles-Privé'] = 0 # Nombre d'élèves filles dans le privé
207 sommes['Moy-Public'] = 0 # Somme des moyennes de filles par établissement dans
208 ↪ le public
209 sommes['Nb-Public'] = 0 # Nombre d'établissements publics considérés (pour faire
210 ↪ la moyenne de moyennes)
211 sommes['Moy-Privé'] = 0 # Somme des moyennes de filles par établissement dans le
212 ↪ privé
213 sommes['Nb-Privé'] = 0 # Nombre d'établissements publics considérés (pour faire
214 ↪ la moyenne de moyennes)
215
216 for i in range(len(table)):
217     # On ne regarde que les lycées qui ont des élèves en seconde (sinon on fera
218     ↪ des divisions par zéro):
219     if table[i]['2ndes_gt'] > 0:
220         if table[i]['secteur'] == "PUBLIC":
221             sommes['2nde-Public'] += table[i]['2ndes_gt']
222             sommes['Filles-Public'] += table[i]['2ndes_gt_filles']
223             sommes['Nb-Public'] += 1 # On ajoute un établissement seulement

```

```

211     sommes['Moy-Public'] += table[i]['2ndes_gt_filles'] /
        ↳ table[i]['2ndes_gt']
212     else:
213         sommes['2nde-Prive'] += table[i]['2ndes_gt']
214         sommes['Filles-Prive'] += table[i]['2ndes_gt_filles']
215         sommes['Nb-Prive'] += 1 # On ajoute un établissement seulement
216         sommes['Moy-Prive'] += table[i]['2ndes_gt_filles'] /
        ↳ table[i]['2ndes_gt']
217
218 # On a maintenant toutes les informations qu'il nous faut - faisons les calculs
        ↳ de nos résultats:
219 # Pourcentages moyens pour commencer
220 PctMoyPub = round(sommes['Filles-Public'] / sommes['2nde-Public'] * 100, 2)
221 PctMoyPrv = round(sommes['Filles-Prive'] / sommes['2nde-Prive'] * 100, 2)
222 PctMoyTot = round((sommes['Filles-Public'] + sommes['Filles-Prive']) /
        ↳ (sommes['2nde-Public'] + sommes['2nde-Prive']) * 100, 2)
223 # Moyennes de pourcentage ensuite
224 MoyMoyPub = round(sommes['Moy-Public'] / sommes['Nb-Public'] * 100, 2)
225 MoyMoyPrv = round(sommes['Moy-Prive'] / sommes['Nb-Prive'] * 100, 2)
226 MoyMoyTot = round((sommes['Moy-Public'] + sommes['Moy-Prive']) /
        ↳ (sommes['Nb-Public'] + sommes['Nb-Prive']) * 100, 2)
227
228 # Il nous reste maintenant à afficher tout ça dans un beau tableau - même si
        ↳ j'ai conscience qu'on n'a
229 # pas encore vu cette syntaxe de print.
230 # En-têtes
231 header = f"{' ':<25} {'Public':<10} {'Privé':<10} {'Total':<10}"
232 # Lignes de données
233 row1 = f"{'Pourcentage moyen':<25} {PctMoyPub:<10} {PctMoyPrv:<10}
        ↳ {PctMoyTot:<10}"
234 row2 = f"{'Moy. de pourcentages':<25} {MoyMoyPub:<10} {MoyMoyPrv:<10}
        ↳ {MoyMoyTot:<10}"
235 # Affichage du tableau
236 print(header)
237 print(row1)
238 print(row2)
239 Sep()
240 # Et c'est fini!
241 print("===== FIN =====")

```