

Entrée [1]:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
```

Entrée [2]:

```
url = "https://www.guide-des-aliments.com/dietetique/aliments_par_nom.html"
```

Entrée [3]:

```
reponse = requests.get(url)
```

Entrée [4]:

```
if reponse.ok:
    print(reponse)
else:
    print("Erreur lors de la connexion")
```

<Response [200]>

Entrée [5]:

```
site = BeautifulSoup(reponse.text)
```

Entrée [6]:

```
site.prettify()
>\n      </td>\n      </tr>\n      <tr height="65">\n      <td height="65">\n      <table border="0" cellpadding="3" cellspacing="2" width="100%">\n      <tr>\n      <td class="titre-article" width="50%">\n      <div align="center">\n      <a href="liste_aliments_lettre_C.html">\n      Aliments C\n      </a>\n      \n      </div>\n      </td>\n      <td class="titre-article">\n      <div align="center">\n      <a href="liste_aliments_lettre_D.html">\n      Aliments D\n      </a>\n      </div>\n      </td>\n      </tr>\n      <tr>\n      <td class="texte-article" valign="top" width="50%">\n      Cao, canard, carotte, cerise, chicorée, cornichon, comté, ...\n      </td>\n      <td class="texte-article" valign="top">\n      Datte, doré, Durian, Doliqu e, Dorade, Dinde, ...\n      </td>\n      </tr>\n      </table>\n      </td>\n      \n      </tr>\n      <tr height="65">\n      <td height="65">\n      <table border="0" cellpadding="3" cellspacing="2" width="100%">\n      <tr>\n      <td class="titre-article" width="50%">\n      <div align="center">\n      <a href="liste_aliments_lettre_E.html">\n      Aliments E\n      </a>\n      </div>\n      </td>\n      <td class="titre-article">\n      <div align="center">\n      <a href="liste_aliments_lettre_F.html">\n      Aliments F\n      </a>\n      </div>\n      </td>\n      </tr>\n      <tr>\n      <td class="texte-article" valign="top" width="50%">\n      Éch alote, escargot, épinard, époisses, eau, esturgeon, ...\n      </td>\n      </tr>\n      </table>\n      </td>\n      </tr>\n      </table>\n      </td>
```

Entrée [7]: `site.find_all("a")`

Out[7]: `[, Aliments, Recettes, Informations-Santé, Contact, Sommaire, Liste des aliments par catégorie, Liste des aliments par famille, Liste des aliments par valeur nutritive, Liste des Recettes par aliment, Liste des Recettes par catégorie, ...]`

Entrée [8]: `type(site.find_all("a")[0])`

Out[8]: `bs4.element.Tag`

Entrée [9]: `print(site.find_all("a")[0].get_text())`

Entrée [10]: `type(site.find_all("a")[0].get_text())`

Out[10]: `str`

```
Entrée [15]: # Liste des balises a qui correspondent aux aliments
a_Aliments = []
for a in site.find_all("a"):
    if "Aliments" in a.get_text():
        a_Aliments.append(a)
a_Aliments
```

```
Out[15]: [Aliments</a>,
Aliments A</a>,
Aliments B</a>,
Aliments C</a>,
Aliments D</a>,
Aliments E</a>,
Aliments F</a>,
Aliments G</a>,
Aliments H</a>,
Aliments I</a>,
Aliments J</a>,
Aliments K</a>,
Aliments L</a>,
Aliments M</a>,
Aliments N</a>,
Aliments O</a>,
Aliments P</a>,
Aliments Q</a>,
Aliments R</a>,
Aliments S</a>,
Aliments T</a>,
Aliments V</a>,
Aliments W</a>,
Aliments Y</a>]
```

```
Entrée [16]: a_Aliments = a_Aliments[1:]
a_Aliments
```

```
Out[16]: [Aliments A</a>,
Aliments B</a>,
Aliments C</a>,
Aliments D</a>,
Aliments E</a>,
Aliments F</a>,
Aliments G</a>,
Aliments H</a>,
Aliments I</a>,
Aliments J</a>,
Aliments K</a>,
Aliments L</a>,
Aliments M</a>,
Aliments N</a>,
Aliments O</a>,
Aliments P</a>,
Aliments Q</a>,
Aliments R</a>,
Aliments S</a>,
Aliments T</a>,
Aliments V</a>,
Aliments W</a>,
Aliments Y</a>]
```

```
Entrée [18]: # Liste qui contient Les Liens des catégories par Lettre
a_Aliments_lettre_liens = []
for a in a_Aliments:
    a_Aliments_lettre_liens.append(str(a.get('href')))
a_Aliments_lettre_liens
```

```
Out[18]: ['liste_aliments_lettre_A.html',
'liste_aliments_lettre_B.html',
'liste_aliments_lettre_C.html',
'liste_aliments_lettre_D.html',
'liste_aliments_lettre_E.html',
'liste_aliments_lettre_F.html',
'liste_aliments_lettre_G.html',
'liste_aliments_lettre_H.html',
'liste_aliments_lettre_I.html',
'liste_aliments_lettre_J.html',
'liste_aliments_lettre_K.html',
'liste_aliments_lettre_L.html',
'liste_aliments_lettre_M.html',
'liste_aliments_lettre_N.html',
'liste_aliments_lettre_O.html',
'liste_aliments_lettre_P.html',
'liste_aliments_lettre_Q.html',
'liste_aliments_lettre_R.html',
'liste_aliments_lettre_S.html',
'liste_aliments_lettre_T.html',
'liste_aliments_lettre_V.html',
'liste_aliments_lettre_W.html',
'liste_aliments_lettre_Y.html']
```

```
Entrée [19]: # On ajoute maintenant Le reste du Lien
a_Aliments_lettre_lien_complet = []
for lien in a_Aliments_lettre_liens:
    a_Aliments_lettre_lien_complet.append("https://www.guide-des-aliments.com/diet
a_Aliments_lettre_lien_complet
```

```
Out[19]: ['https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_A.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_B.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_C.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_D.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_E.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_F.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_G.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_H.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_I.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_J.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_K.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_L.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_M.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_N.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_O.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_P.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_Q.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_R.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_S.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_T.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_V.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_W.html',
'https://www.guide-des-aliments.com/dietetique/liste_aliments_lettre_Y.html']
```

```
Entrée [20]: # Etape 2
# Test sur le premier Lien
reponse = requests.get(a_Aliments_lettre_lien_complet[0])
if reponse.ok:
    print(reponse)
else:
    print("Erreur lors de la connexion")
```

<Response [200]>

```
Entrée [21]: site = BeautifulSoup(reponse.text)
site.prettify()
site.find_all("a")

<a href="fiche_agar_499.html"></a>,
<a href="fiche_agar_499.html">Agar</a>,
<a href="fiche_agave_2.html"></a>,
<a href="fiche_agave_2.html">Agave</a>,
<a href="fiche_agneau_3.html"></a>,
<a href="fiche_agneau_3.html">Agneau</a>,
<a href="fiche_ail_4.html"></a>,
<a href="fiche_ail_4.html">Ail</a>,
<a href="fiche_airelle_5.html"></a>,
<a href="fiche_airelle_5.html">Airelle</a>,
<a href="fiche_alcool_6.html"></a>,
<a href="fiche_alcool_6.html">Alcool</a>,
<a href="fiche_algue_7.html"></a>
```

```
Entrée [27]: # premier filtre
a_fiches = []
for a in site.find_all("a"):
    if "fiche" in str(a.get('href')):
        a_fiches.append(a)
a_fiches
```

```
Out[27]: [<a href="fiche_abricot_1.html"></a>,
  <a href="fiche_abricot_1.html">Abricot</a>,
  <a href="fiche_achigan_498.html"></a>,
  <a href="fiche_achigan_498.html">Achigan</a>,
  <a href="fiche_agar_499.html"></a>,
  <a href="fiche_agar_499.html">Agar</a>,
  <a href="fiche_agave_2.html"></a>,
  <a href="fiche_agave_2.html">Agave</a>,
  <a href="fiche_agneau_3.html"></a>,
  <a href="fiche_agneau_3.html">Agneau</a>,
  <a href="fiche_ail_4.html"></a>,
  <a href="fiche_ail_4.html">Ail</a>,
  <a href="fiche_airelle_5.html"></a>,
  <a href="fiche_airelle_5.html">Airelle</a>,
  <a href="fiche_alcool_6.html"></a>,
  <a href="fiche_alcool_6.html">Alcool</a>,
  <a href="fiche_algue_7.html"></a>,
  <a href="fiche_algue_7.html">Algue</a>,
  <a href="fiche_alkekenge_8.html"></a>,
  <a href="fiche_alkekenge_8.html">Alkékenge</a>,
  <a href="fiche_alose_9.html"></a>,
  <a href="fiche_alose_9.html">Alose</a>,
  <a href="fiche_amande_10.html"></a>,
  <a href="fiche_amande_10.html">Amande</a>,
  <a href="fiche_amarante_11.html"></a>,
  <a href="fiche_amarante_11.html">Amarante</a>,
  <a href="fiche_ananas_12.html"></a>,
  <a href="fiche_ananas_12.html">Ananas</a>,
  <a href="fiche_anchois_13.html"></a>,
  <a href="fiche_anchois_13.html">Anchois</a>,
  <a href="fiche_andouille_500.html"></a>,
  <a href="fiche_andouille_500.html">Andouille</a>,
  <a href="fiche_andouillette_501.html"></a>,
  <a href="fiche_andouillette_501.html">Andouillette</a>,
  <a href="fiche_aneth_14.html"></a>,
  <a href="fiche_aneth_14.html">Aneth</a>,
  <a href="fiche_angelique_502.html"></a>,  
<a href="fiche_angelique_502.html">Angélique</a>,  
<a href="fiche_anguille_16.html"></a>,  
<a href="fiche_anguille_16.html">Anguille</a>,  
<a href="fiche_anguille-de-roches_15.html"></a>,  
<a href="fiche_anguille-de-roches_15.html">Anguille de roches</a>,  
<a href="fiche_anis_17.html"></a>,  
<a href="fiche_anis_17.html">Anis</a>,  
<a href="fiche_antilope_18.html"></a>,  
<a href="fiche_antilope_18.html">Antilope</a>,  
<a href="fiche_arachide_19.html"></a>,  
<a href="fiche_arachide_19.html">Arachide</a>,  
<a href="fiche_aramé_503.html"></a>,  
<a href="fiche_aramé_503.html">Aramé</a>,  
<a href="fiche_artichaut_20.html"></a>,  
<a href="fiche_artichaut_20.html">Artichaut</a>,  
<a href="fiche_ascidie_21.html"></a>,  
<a href="fiche_ascidie_21.html">Ascidie</a>,  
<a href="fiche_asperge_22.html"></a>,  
<a href="fiche_asperge_22.html">Asperge</a>,  
<a href="fiche_aubepine_23.html"></a>,  
<a href="fiche_aubepine_23.html">Aubépine</a>,  
<a href="fiche_aubergine_24.html"></a>,  
<a href="fiche_aubergine_24.html">Aubergine</a>,  
<a href="fiche_autruche_25.html"></a>,  
<a href="fiche_autruche_25.html">Autruche</a>,  
<a href="fiche_avocat_26.html"></a>,  
<a href="fiche_avocat_26.html">Avocat</a>,  
<a href="fiche_avoine_27.html"></a>,  
<a href="fiche_avoine_27.html">Avoine</a>,  
<a href="fiche_azerole_28.html"></a>,  
<a href="fiche_azerole_28.html">Azerole</a>]
```

```
Entrée [35]: # récupération des noms
liens = []
for a in a_fiches:
    nom= a.get_text()
    if len(nom) > 1:
        liens.append(a.get("href"))
liens
```

```
Out[35]: ['fiche_abricot_1.html',
'fiche_achigan_498.html',
'fiche_agar_499.html',
'fiche_agave_2.html',
'fiche_agneau_3.html',
'fiche_ail_4.html',
'fiche_airelle_5.html',
'fiche_alcool_6.html',
'fiche_algue_7.html',
'fiche_alkekenge_8.html',
'fiche_alose_9.html',
'fiche_amande_10.html',
'fiche_amarante_11.html',
'fiche_ananas_12.html',
'fiche_anchois_13.html',
'fiche_andouille_500.html',
'fiche_andouillette_501.html',
'fiche_aneth_14.html',
'fiche_angelique_502.html',
'fiche_anguille_16.html',
'fiche_anguille-de-roches_15.html',
'fiche_anis_17.html',
'fiche_antilope_18.html',
'fiche_arachide_19.html',
'fiche_aramé_503.html',
'fiche_artichaut_20.html',
'fiche_ascidie_21.html',
'fiche_asperge_22.html',
'fiche_aubepine_23.html',
'fiche_aubergine_24.html',
'fiche_autruche_25.html',
'fiche_avocat_26.html',
'fiche_avoine_27.html',
'fiche_azerole_28.html']
```


Entrée [37]: *# Application du modèle de scrap à tous les Liens récoltés précédement*

```
liens = []
# Parcourir tous Les Liens
for url in a_Aliments_lettre_lien_complet:
    reponse = requests.get(url)
    if reponse.ok:
        print(reponse)
        site = BeautifulSoup(reponse.text)
        site.prettify()
        site.find_all("a")
        # premier filtre
        a_fiches = []
        for a in site.find_all("a"):
            if "fiche" in str(a.get('href')):
                a_fiches.append(a)
        # Permet ne pas faire de doublons de Les Liens
        for a in a_fiches:
            nom = a.get_text()
            if len(nom) > 1:
                liens.append(a.get("href"))
        else:
            print("Erreur lors de la connexion")

print(f"Nombre de liens : {len(liens)}")
liens
```

```
'fiche_dore-noir_178.html',
'fiche_durian_179.html',
'fiche_echalote_29.html',
'fiche_ecrevisse_30.html',
'fiche_endive_497.html',
'fiche_eperlan_181.html',
'fiche_epinard_38.html',
'fiche_erable_39.html',
'fiche_escargot_182.html',
'fiche_espadon_183.html',
'fiche_estragon_184.html',
'fiche_esturgeon_185.html',
'fiche_faine_186.html',
'fiche_faisan_187.html',
'fiche_feijoa_188.html',
'fiche_fenouil_189.html',
'fiche_fenugrec_190.html',
'fiche_feve_544.html',
'fiche_figue_194.html',
'fiche_figue-de-barbarie_193.html',
'fiche_figue-de-barbarie_193.html',
```

```
Entrée [41]: # On ajoute maintenant le reste du lien
a_Aliments_liens_complets = []
for lien in liens:
    a_Aliments_liens_complets.append("https://www.guide-des-aliments.com/dietetique/lien/lien.html")
print(f"Nombre de liens : {len(a_Aliments_liens_complets)}")
a_Aliments_liens_complets
```

```
Out[41]: ['https://www.guide-des-aliments.com/dietetique/fiche_abricot_1.html',
'https://www.guide-des-aliments.com/dietetique/fiche_achigan_498.html',
'https://www.guide-des-aliments.com/dietetique/fiche_agar_499.html',
'https://www.guide-des-aliments.com/dietetique/fiche_agave_2.html',
'https://www.guide-des-aliments.com/dietetique/fiche_agneau_3.html',
'https://www.guide-des-aliments.com/dietetique/fiche_ail_4.html',
'https://www.guide-des-aliments.com/dietetique/fiche_airelle_5.html',
'https://www.guide-des-aliments.com/dietetique/fiche_alcool_6.html',
'https://www.guide-des-aliments.com/dietetique/fiche_algue_7.html',
'https://www.guide-des-aliments.com/dietetique/fiche_alkekonge_8.html',
'https://www.guide-des-aliments.com/dietetique/fiche_alose_9.html',
'https://www.guide-des-aliments.com/dietetique/fiche_amande_10.html',
'https://www.guide-des-aliments.com/dietetique/fiche_amarante_11.html',
'https://www.guide-des-aliments.com/dietetique/fiche_ananas_12.html',
```

```
Entrée [42]: # étape 3
# Test sur le premier lien

reponse = requests.get(a_Aliments_liens_complets[0])
if reponse.ok:
    print(reponse)
else:
    print("Erreur lors de la connexion")
```

<Response [200]>

```
Entrée [43]: site = BeautifulSoup(reponse.text)
site.prettify()
site.find_all("a")
```

```
Out[43]: [<a href="https://www.guide-des-aliments.com"></a>,
  <a href="aliments_par_nom.html">
Aliments</a>,
  <a href="recettes_par_nom.html">
Recettes</a>,
  <a href="Information/sommaire.html">Informations-Santé</a>,
  <a href="Contact/index.php">
Contact</a>,
  <a href="index.php">Sommaire</a>,
  <a href="aliments_par_nom.html">Liste des aliments par nom</a>,
  <a href="fruits_3.html">Fruits</a>,
  <a href="liste_aliments_lettre_A.html">Aliment A</a>,
  <a href="fiche_nutrition_abricot_1.html">Calories, vitamines de l'abricot</a>,
  <a href="Article_1_abricot_sec_1-31.html">L'abricot sec</a>,
  <a href="Article_1_abricot_en_boite_1-32.html">L'abricot en boîte</a>,
  <a href="aliments_par_categorie.html">Liste des aliments par catégorie</a>,
  <a href="aliments_par_famille.html">Liste des aliments par famille</a>,
  <a href="aliments_par_valeur_nutritive.html">Liste des aliments par valeur nutritive</a>,
  <a href="recettes_par_aliments.html">Liste des Recettes par aliment</a>,
  <a href="recettes_par_categorie.html">Liste des Recettes par catégorie</a>,
  <a href="recettes_par_valeur_nutritive.html">Liste des Recettes par valeur nutritive</a>,
  <a href="liste_recettes_aliment_abricot_1_1.html"><b><font size="3">Recettes à
base
d'abricot</font></b></a>,
  <a href="recette_gateau_aux_abricots_secs_268.html">Recette du gâteau aux abri
cots secs</a>,
  <a href="recette_marmelade_d_abricots_et_rhubarbe_492.html">Recette de la marme
lade d'abricots et rhubarbe</a>,
  <a href="recette_macedoine_de_fruits_au_sirop_486.html">Recette de la macédoine
de fruits au sirop</a>,
  <a href="https://www.guide-des-aliments.com/dietetique/Information/Micro-organi
smes/AA-Micro-organisme.html">Les micro-organismes et les aliments</a>,
  <a href="https://www.guide-des-aliments.com/dietetique/Information/Micro-organi
smes/BA-Contamination-aliments.html">Contamination et la détérioration</a>,
  <a href="https://www.guide-des-aliments.com/dietetique/Information/Micro-organi
smes/DA-Toxi-Infections.html">Toxi-infection et intoxication</a>,
  <a href="https://www.guide-des-aliments.com/dietetique/Information/Micro-organi
smes/CA-Contributions.html">Contribution des micro-organismes</a>,
  <a href="https://omega.xetaz.net/adclick.php?n=a4faa74e" target="_blank"></a>,
  <a href="fiche_nutrition_abricot_1.html">Calories et vitamines de l'abricot</a
>,
  <a href="https://omega.xetaz.net/adclick.php?n=a4d27983" target="_blank"></a>]
```

```
Entrée [45]: for a in site.find_all('a'):
              if "Calories," in a.get_text():
                  print(a.get('href'))
```

fiche_nutrition_abricot_1.html

```

Entrée [48]: # Application du modèle pour tous les liens
liens_fiches_nutrition = []
# Parcourir tous les liens
for url in a_Aliments_liens_complets:
    reponse = requests.get(url)
    if reponse.ok:
        print(reponse)
        site = BeautifulSoup(reponse.text)
        site.prettify()
        site.find_all("a")
        # Récupérer le lien vers la fiche
        for a in site.find_all('a'):
            if "Calories," in a.get_text():
                print("Récupération du lien de la fiche...")
                liens_fiches_nutrition.append(str(a.get('href')))
        else:
            print("Erreur lors de la connexion")

print(f"Nombre de liens : {len(liens_fiches_nutrition)}")

```

```

<Response [200]>
Récupération du lien de la fiche...
<Response [200]>
Récupération du lien de la fiche...
<Response [200]>
Récupération du lien de la fiche...
<Response [200]>
Récupération du lien de la fiche...
<Response [200]>
Récupération du lien de la fiche...
<Response [200]>
Récupération du lien de la fiche...
<Response [200]>
Récupération du lien de la fiche...
<Response [200]>
Récupération du lien de la fiche...
<Response [200]>
Récupération du lien de la fiche...
<Response [200]>

```

```

Entrée [49]: liens_fiches_nutrition
'fiche_nutrition_anis_17.html',
'fiche_nutrition_antilope_18.html',
'fiche_nutrition_arachide_19.html',
'fiche_nutrition_arames_503.html',
'fiche_nutrition_artichaut_20.html',
'fiche_nutrition_ascidie_21.html',
'fiche_nutrition_asperge_22.html',
'fiche_nutrition_aubepine_23.html',
'fiche_nutrition_aubergine_24.html',
'fiche_nutrition_autruche_25.html',
'fiche_nutrition_avocat_26.html',
'fiche_nutrition_avoine_27.html',
'fiche_nutrition_azerole_28.html',
'fiche_nutrition_babassu_40.html',
'fiche_nutrition_babeurre_504.html',
'fiche_nutrition_baguette_de_pain_538.html',
'fiche_nutrition_baie_de_genievre_505.html',
'fiche_nutrition_baleine_41.html',
'fiche_nutrition_bambou_506.html',
'fiche_nutrition_banane_42.html',

```

Entrée [50]: *# On ajoute maintenant le reste du lien*

```
liens_fiches_nutrition_complets = []
for lien in liens_fiches_nutrition:
    liens_fiches_nutrition_complets.append("https://www.guide-des-aliments.com/die")
print(f"Nombre de liens : {len(liens_fiches_nutrition_complets)}")
liens_fiches_nutrition_complets
```

Nombre de liens : 387

Entrée [69]: *# Test sur un lien*

```
url = liens_fiches_nutrition_complets[0]
reponse = requests.get(url, )
reponse.encoding = reponse.apparent_encoding
reponse
```

Out[69]: <Response [200]>

```
Entrée [70]: site = BeautifulSoup(reponse.text)
site.prettify()
site.find_all("a")
```

```
Out[70]: [<a href="https://www.guide-des-aliments.com"></a>,
  <a href="aliments_par_nom.html">

Aliments</a>,
  <a href="recettes_par_nom.html">

Recettes</a>,
  <a href="Information/sommaire.html">Informations-Santé</a>,
  <a href="Contact/index.php">

Contact</a>,
  <a href="index.php">Sommaire</a>,
  <a href="aliments_par_nom.html">Liste des aliments par nom</a>,
  <a href="fruits_3.html">Fruits</a>,
  <a href="liste_aliments_lettre_A.html">Aliment A</a>,
  <a href="fiche_abricot_1.html">Abricot</a>,
  <a href="Article_1_abricot_sec_1-31.html">L'abricot sec</a>,
  <a href="Article_1_abricot_en_boite_1-32.html">L'abricot en boîte</a>,
  <a href="aliments_par_categorie.html">Liste des aliments par catégorie</a>,
  <a href="aliments_par_famille.html">Liste des aliments par famille</a>,
  <a href="aliments_par_valeur_nutritive.html">Liste des aliments par valeur nutr
itive</a>,
  <a href="recettes_par_aliments.html">Liste des Recettes par aliment</a>,
  <a href="recettes_par_categorie.html">Liste des Recettes par catégorie</a>,
  <a href="recettes_par_valeur_nutritive.html">Liste des Recettes par valeur nutr
itive</a>,
  <a href="liste_recettes_aliment_abricot_1_1.html"><b><font size="3">Recettes à
base
d'abricot</font></b></a>,
  <a href="recette_chaussons_aux_fruits_355.html">Recette des chaussons aux fruit
s</a>,
  <a href="recette_marmelade_d_abricots_et_rhubarbe_492.html">Recette de la marme
lade d'abricots et rhubarbe</a>,
  <a href="recette_escalopes_de_volaille_aux_abricots_222.html">Recette des escal
opes de volaille aux abricots</a>,
  <a href="https://www.guide-des-aliments.com/dietetique/Information/Micro-organi
smes/AA-Micro-organisme.html">Les micro-organismes et les aliments</a>,
  <a href="https://www.guide-des-aliments.com/dietetique/Information/Micro-organi
smes/BA-Contamination-aliments.html">Contamination et la détérioration</a>,
  <a href="https://www.guide-des-aliments.com/dietetique/Information/Micro-organi
smes/DA-Toxi-Infections.html">Toxi-infection et intoxication</a>,
  <a href="https://www.guide-des-aliments.com/dietetique/Information/Micro-organi
smes/CA-Contributions.html">Contribution des micro-organismes</a>,
  <a href="https://omega.xetaz.net/adclick.php?n=a4faa74e" target="_blank"></a>,
  <a href="detail_nutrition_abricot-huile-de-noyaux-d-abricot_441.html">Abricot,
huile de noyaux d'abricot</a>,
  <a href="detail_nutrition_abricot-cru_1498.html">Abricot, cru</a>,
  <a href="detail_nutrition_abricot-en-conserve-a-l-eau_1499.html">Abricot, en co
nserve, à l'eau</a>,
  <a href="detail_nutrition_abricot-en-conserve-au-sirop-leger_1503.html">Abrico
t, en conserve, au sirop léger</a>,
  <a href="detail_nutrition_abricot-en-conserve-au-sirop-epais_1504.html">Abrico
t, en conserve, au sirop épais</a>,
  <a href="detail_nutrition_abricot-congele_1510.html">Abricot, congelé</a>,
  <a href="detail_nutrition_abricot-des-antilles-cru_1602.html">Abricot, des Anti
lles, cru</a>],
```

```

<a href="detail_nutrition_abricot-nectar_1694.html">Abricot, nectar</a>,
<a href="detail_nutrition_abricot-sec-cru_1702.html">Abricot, sec, cru</a>,
<a href="detail_nutrition_abricot-sec-cuit_1703.html">Abricot, sec, cuit</a>,
<a href="detail_nutrition_abricot-confiture-et-en-conserve_4379.html">Abricot,
confiture et en conserve</a>,
<a href="detail_nutrition_abricot-confit_4395.html">Abricot, confit</a>,
<a href="detail_nutrition_abricot-en-conserve-au-sirop-epais-egoutte_5395.htm
l">Abricot, en conserve, au sirop épais, égoutté</a>,
<a href="https://omega.xetaz.net/adclick.php?n=a4d27983" target="_blank"></a>]

```

```

Entrée [71]: for a in site.find_all('a'):
               if "detail_nutrition" in str(a.get('href')):
                   print(a.get_text())

```

```

Abricot, huile de noyaux d'abricot
Abricot, cru
Abricot, en conserve, à l'eau
Abricot, en conserve, au sirop léger
Abricot, en conserve, au sirop épais
Abricot, congelé
Abricot, des Antilles, cru
Abricot, nectar
Abricot, sec, cru
Abricot, sec, cuit
Abricot, confiture et en conserve
Abricot, confit
Abricot, en conserve, au sirop épais, égoutté

```

```
Entrée [82]: # Application du modèle pour tous les liens
noms = []
liens_aliments = []
i = 1
# Parcourir tous les liens
for url in liens_fiches_nutrition_complets:
    reponse = requests.get(url)
    reponse.encoding = reponse.apparent_encoding
    if reponse.ok:
        print(f"Lien{i}/{len(liens_fiches_nutrition_complets)}")
        site = BeautifulSoup(reponse.text)
        site.prettify()
        site.find_all("a")
        # Récupérer le lien vers la fiche
        for a in site.find_all('a'):
            if "detail_nutrition" in str(a.get('href')):
                noms.append(a.get_text())
                liens_aliments.append("https://www.guide-des-aliments.com/dietetico
    else:
        print("Erreur lors de la connexion")
    i += 1
print(f"Nombre d'aliements: {len(liens_aliments)}")
Lien361/387
Lien362/387
Lien363/387
Lien364/387
Lien365/387
Lien366/387
Lien367/387
Lien368/387
Lien369/387
Lien370/387
Lien371/387
Lien372/387
Lien373/387
Lien374/387
Lien375/387
Lien376/387
Lien377/387
Lien378/387
Lien379/387
Lien380/387
```


Entrée [83]: liens_aliments[2300:]

Out[83]: ['https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-cervelle-cru_3603.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-cervelle-braise_3604.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-cervelle-saute_3605.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-coeur-cru_3608.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-coeur-braise_3609.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-rognon-cru_3612.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-rognon-braise_3613.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-foie-cru_3617.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-foie-braise_3618.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-foie-saute_3619.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-poumon-cru_3622.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-poumon-braise_3623.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-pancreas-cru_3627.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-pancreas-braise_3628.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-rate-cru_3631.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-rate-braise_3632.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-ris-cru_3633.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-ris-braise_3634.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-langue-cru_3637.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_veau-langue-braise_3638.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vergerette-du-canada-feuille-crue_4865.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-de-dessert-liqueux-18-de-volume-d-alcool_2848.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-de-table-115-de-volume-d-alcool_2849.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-de-table-rouge-115-de-volume-d-alcool_2850.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-de-table-rose-115-de-volume-d-alcool_2851.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-de-table-blanc-115-de-volume-d-alcool_2852.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-de-dessert-sec-18-de-volume-d-alcool_2937.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-ordinaire_2978.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-sans-alcool_5294.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-de-cuisine_5551.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vin-de-table-

leger_5552.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vinaigre-de-cidre_13.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vinaigre-blanc_14.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vinaigre-de-vin_6195.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vinaigre-balsamique_6196.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vivaneau-crues_3066.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_vivaneau-cuit_3067.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wakame_crue_2321.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wapiti-viande-cru_3581.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wapiti-viande-cuit_3582.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wapiti-hache-cru_5723.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wapiti-hache-cuit_5724.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wapiti-longe-cuit_5725.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wapiti-ronde-cuit_5726.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wapiti-filet-cuit_5727.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wasabi-condiment_4868.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_wonton-pate_4001.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_worcestershire-condiment_1133.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt--nature--lait-ecreme_81.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt--nature--lait-demi-ecreme_141.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt-nature-0_142.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt--nature--lait-entier_154.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt-nature-brasse_155.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt-nature-a-boire_156.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt-nature-glace_4292.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt__petit-suisse_a_40_m-g_6300.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt__petit-suisse_a_60_m-g_6301.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt__fromage_blanc_battu_a_0_mg_6401.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt__fromage_blanc_battu_a_20_mg_6402.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt__fromage_blanc_battu_a_30_mg_6403.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt__fromage_blanc_battu_a_40_mg_6404.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt__fromage_blanc_a_0_mg_6405.html',
'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yaourt__froma

```
ge_blanc_a_40_mg_6406.html',  
  'https://www.guide-des-aliments.com/dietetique/detail_nutrition_yogourt__natu  
re_6428.html']
```

Entrée [84]: `print(f"Nombre d'aliments: {len(noms)}")`

Nombre d'aliments: 2364

Entrée [85]: noms[2300:]

```
Out[85]: ['Veau, cervelle, cru',
'Veau, cervelle, braisé',
'Veau, cervelle, sauté',
'Veau, coeur, cru',
'Veau, coeur, braisé',
'Veau, rognon, cru',
'Veau, rognon, braisé',
'Veau, foie, cru',
'Veau, foie, braisé',
'Veau, foie, sauté',
'Veau, poumon, cru',
'Veau, poumon, braisé',
'Veau, pancréas, cru',
'Veau, pancréas, braisé',
'Veau, rate, cru',
'Veau, rate, braisé',
'Veau, ris, cru',
'Veau, ris, braisé',
'Veau, langue, cru',
'Veau, langue, braisé',
'Vergerette du Canada, feuille, crue',
'Vin, de dessert, liquoreux, 18% de volume d'alcool",
'Vin, de table, 11,5% de volume d'alcool",
'Vin, de table, rouge, 11,5% de volume d'alcool",
'Vin, de table, rosé, 11,5% de volume d'alcool",
'Vin, de table, blanc, 11,5% de volume d'alcool",
'Vin, de dessert, sec, 18% de volume d'alcool",
'Vin, ordinaire',
'Vin, sans alcool',
'Vin, de cuisine',
'Vin, de table, léger',
'Vinaigre, de cidre',
'Vinaigre, blanc',
'Vinaigre, de vin',
'Vinaigre, balsamique',
'Vivaneau, crues',
'Vivaneau, cuit',
'Wakamé, crue',
'Wapiti, viande, cru',
'Wapiti, viande, cuit',
'Wapiti, haché, cru',
'Wapiti, haché, cuit',
'Wapiti, longue, cuit',
'Wapiti, ronde, cuit',
'Wapiti, filet, cuit',
'Wasabi, condiment',
'Wonton, pâte',
'Worcestershire, condiment',
'Yaourt, nature, lait écrémé',
'Yaourt, nature, lait demi-écrémé',
'Yaourt, nature, 0%',
'Yaourt, nature, lait entier',
'Yaourt, nature, brassé',
'Yaourt, nature, à boire',
'Yaourt, nature, glacé',
'Yaourt, petit-suisse à 40 % M.G',
'Yaourt, petit-suisse à 60 % M.G',
'Yaourt, fromage blanc battu à 0% MG',
'Yaourt, fromage blanc battu à 20% MG',
'Yaourt, fromage blanc battu à 30% MG',
'Yaourt, fromage blanc battu à 40% MG',
```

```
'Yaourt, fromage blanc à 0% MG',  
'Yaourt, fromage blanc à 40% MG',  
'Yogourt, nature']
```

```
Entrée [86]: # étape 4  
# Test sur le premier lien  
reponse = requests.get(liens_aliments[0])  
reponse.encoding = reponse.apparent_encoding  
reponse
```

Out[86]: <Response [200]>

```
Entrée [88]: site = BeautifulSoup(reponse.text)  
site.prettify()  
site.find_all("td")  
  
<tr height="300">  
<td align="center" class="texte-article" height="300" valign="middle">  
<div align="center">  
<script language="JavaScript" src="https://omega.xetaz.net/adx.js" type="text/javascript"></script>  
<script language="JavaScript" type="text/javascript">  
<!--  
    if (!document.phpAds_used) document.phpAds_used = ',';  
    phpAds_random = new String (Math.random()); phpAds_random = phpAds_random.  
substring(2,11);  
  
    document.write ("<" + "script language='JavaScript' type='text/javascript'  
src='");  
    document.write ("https://omega.xetaz.net/adjs.php?n=" + phpAds_random);  
    document.write ("&what=zone:28");  
    document.write ("&exclude=" + document.phpAds_used);  
    if (document.referrer)  
        document.write ("&referer=" + escape(document.referrer));  
    document.write ("'><" + "/script>");  
    //-->
```

```
Entrée [96]: def is_a_number(string):  
    try:  
        float(string)  
        return True  
    except:  
        return False
```

```
Entrée [97]: quantites = []  
for td in site.find_all("td"):  
    texte = td.get_text().split(" ")  
  
    if len(texte) == 2 and is_a_number(texte[0]):  
        print(td.get_text())
```

884 kCal

100.000 g

6.300 g

60.000 g

29.300 g

0.000 g

0.000 g

0.000 g

0.000 g

0.000 mg

0.000 µg

0.000 mg

0.000 mg

0.000 mg

0.000 mg

0.000 mg

0.000 µg

0.000 µg

0.000 mg

0.000 µg

4.000 mg

0.000 mg

0.000 mg

0.000 mg

0.000 mg

0.000 mg

0.000 mg

0.000 mg

0.000 µg

```
Entrée [99]: # Application du modèle pour tous les liens
quantites = []
i = 1
# Parcourir tous les liens
for url in liens_aliments:
    reponse = requests.get(url)
    reponse.encoding = reponse.apparent_encoding
    if reponse.ok:
        print(f"Lien{i}/{len(liens_aliments)}")
        site = BeautifulSoup(reponse.text)
        site.prettify()
        quantites_produit = []
        for td in site.find_all("td"):
            texte = td.get_text().split(" ")
            if len(texte) == 2 and is_a_number(texte[0]):
                quantites_produit.append(td.get_text())
        else:
            print("Erreur lors de la connexion")
        quantites.append(quantites_produit)
    i += 1
Lien2218/2364
Lien2219/2364
Lien2220/2364
Lien2221/2364
Lien2222/2364
Lien2223/2364
Lien2224/2364
Lien2225/2364
Lien2226/2364
Lien2227/2364
Lien2228/2364
Lien2229/2364
Lien2230/2364
Lien2231/2364
Lien2232/2364
Lien2233/2364
Lien2234/2364
Lien2235/2364
Lien2236/2364
Lien2237/2364
```

```
Entrée [100]: len(quantites)
```

```
Out[100]: 2364
```


Entrée [103]: quantites[2000]

```
Out[103]: ['\n371 kCal\n',
            '\n0.560 g\n',
            '\n0.151 g\n',
            '\n0.173 g\n',
            '\n0.149 g\n',
            '\n81.720 g\n',
            '\n1.700 g\n',
            '\n6.790 g\n',
            '\n10.160 g\n',
            '\n0.000 mg\n',
            '\n0.000 µg\n',
            '\n0.100 mg\n',
            '\n0.070 mg\n',
            '\n3.632 mg\n',
            '\n1.133 mg\n',
            '\n0.350 mg\n',
            '\n17.000 µg\n',
            '\n0.000 µg\n',
            '\n0.000 mg\n',
            '\n0.000 µg\n',
            '\n0.090 mg\n',
            '\n5.000 mg\n',
            '\n60.000 mg\n',
            '\n136.000 mg\n',
            '\n1.500 mg\n',
            '\n31.000 mg\n',
            '\n120.000 mg\n',
            '\n0.960 mg\n',
            '\n0.191 mg\n',
            '\n23.000 µg\n']
```

```
Entrée [105]: quantites_clean = []
               for q in quantites:
                   quantites_produit_clean = []
                   for q2 in q:
                       quantites_produit_clean.append(q2.replace("\n", ""))
                   quantites_clean.append(quantites_produit_clean)
```

Entrée [134]: quantites_clean[2000][1]

```
Out[134]: '0.560 g'
```

```
Entrée [148]: def concat_listes_indice(liste, indice):
    liste_concat = []
    for i in range(len(liste)):
        try:
            liste_concat.append(liste[i][indice])
        except:
            liste_concat.append("? g")
    return liste_concat

data = {
    "Nom" : noms,
    "Calories" : concat_listes_indice(quantites_clean, 0),
    "Lipides" : concat_listes_indice(quantites_clean, 1),
    "Glucides" : concat_listes_indice(quantites_clean, 5),
    "Protéines" : concat_listes_indice(quantites_clean, 6),
}
```

```
Entrée [149]: df = pd.DataFrame.from_dict(data)
df
```

Out[149]:

	Nom	Calories	Lipides	Glucides	Protéines
0	Abricot, huile de noyaux d'abricot	884 kCal	100.000 g	0.000 g	0.000 g
1	Abricot, cru	48 kCal	0.390 g	11.120 g	2.000 g
2	Abricot, en conserve, à l'eau	27 kCal	0.160 g	6.390 g	1.600 g
3	Abricot, en conserve, au sirop léger	63 kCal	0.050 g	16.490 g	1.600 g
4	Abricot, en conserve, au sirop épais	83 kCal	0.080 g	21.470 g	1.600 g
...
2359	Yaourt, fromage blanc battu à 30% MG	105 kCal	6.100 g	3.300 g	9.000 g
2360	Yaourt, fromage blanc battu à 40% MG	113 kCal	7.700 g	3.600 g	7.300 g
2361	Yaourt, fromage blanc à 0% MG	42 kCal	0.070 g	4.900 g	5.000 g
2362	Yaourt, fromage blanc à 40% MG	91 kCal	6.700 g	3.700 g	4.200 g
2363	Yogourt, nature	66 kCal	3.500 g	4.500 g	4.000 g

2364 rows × 5 columns

```
Entrée [151]: df.to_excel("Aliments.xlsx", index=False)
df.to_csv("Aliments.csv", index=False)
```

```
Entrée [152]: url = "https://www.google.com/search?q=Melon+d%27hiver+chinois%2C+cru&tbm=isch&ve

reponse = requests.get(url)
reponse
```

Out[152]: <Response [200]>

```
Entrée [153]: site = BeautifulSoup(reponse.text)
              site.prettify()
              site.find_all("img")
```

```
Out[153]: [,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
]
```

```
Entrée [ ]: import Google-Image-Scraper/GoogleImageScraper.py
```