

DataEngineer Workshop

Cas d'études : [Pétit aperçu du monde de la data engineering](#)

stive.watat@facsciences-uy1.cm

Rédigé par : stive watat, cycle master informatique, science de données
Université de Yaoundé 1

Tous le fichiers et codes relatifs a ce workshop sont disponible via mon github
<https://github.com/WATAT-YONDEP-STIVE-KEVIN>

February 17, 2025

Contents

| | | |
|-------|---|----|
| 0.1 | Définition Général du projet et exigences | 4 |
| 0.1.1 | Définition du projet | 4 |
| 0.1.2 | Exigences | 4 |
| 0.2 | Conception de la base de données | 5 |
| 0.2.1 | Définition des besoins | 5 |
| 0.2.2 | Création des différentes bases de données et tables | 5 |
| 0.2.3 | Création des bases de données | 5 |
| 0.2.4 | Création des tables | 7 |
| 0.3 | quelques scénarios de la vie réelle pour un DataEngineer | 11 |
| 0.3.1 | Trouver tous les patients avec un diagnostic spécifique | 11 |
| 0.3.2 | Lister tous les rendez-vous pour un medecin spécifique | 11 |
| 0.3.3 | Trouver les médecins qui n'ont aucun rendez-vous | 12 |
| 0.4 | Sauvegarde Journalier des données de CamairCo et de l'hôpital central | 13 |
| 0.4.1 | Mise en situation | 13 |
| 0.4.2 | Cas de CamairCo | 13 |
| 0.5 | Collection des données | 14 |
| 0.5.1 | Collection des données à partir des fichiers | 14 |
| 0.5.2 | Collection des données à partir du web | 15 |
| 0.5.3 | Cas pratiques: scraper les données météo | 16 |
| 0.5.4 | Collection des données à partir des bases de données | 17 |
| 0.5.5 | Collection des données en utilisant les API | 18 |
| 0.6 | Conclusion | 19 |

0.1 Définition Général du projet et exigences

0.1.1 Définition du projet

Comment gérer, comprendre, et analyser les données dans un monde où les données sont de plus en plus grandissantes? Et bien bienvenue dans le monde des [Data Engineer](#), le but premier de ce petit projet est de familiariser les apprenants avec le monde de l'ingénierie des données encore dit Data Engineering en aidant les apprenants à comprendre les points importants et tout ce qu'ils doivent savoir sur l'ingénierie des données, des techniques de bases aux techniques avancées

0.1.2 Exigences

Les exigences liées à ce projet sont:

- Connaissance en SQL
- connaissance en SGBD : postgresSQL
- connaissance de base du langage python et utilisation de ses bibliothèques
- connaissance des structures de base des sites web

0.2 Conception de la base de données

0.2.1 Définition des besoins

Pour besoin de réalisation de ce tutoriel, considérons deux bases de données :
une base de données de l'hôpital central de Yaoundé appelé **HospitalDB** constitué de :

- **Patients**, qui contient les informations sur les patients
- **Doctors**, qui contiennent les informations sur les différents docteurs de l'hôpital
- **Appointments**, contenant les différents rendez-vous du personnel
- **Medications**, contenant les différents types de médicaments que peut offrir l'hôpital

et une autre base de données de la compagnie aérienne CamairCo appelé **AirlineDB** constitué des tables suivantes:

- **Flights**, qui contient les informations sur les vols
- **Passengers**, qui contiennent les informations sur les différents passagers
- **Crew**, contenant les différents équipements
- **Booking**, contenant les différentes réservations

0.2.2 Création des différentes bases de données et tables

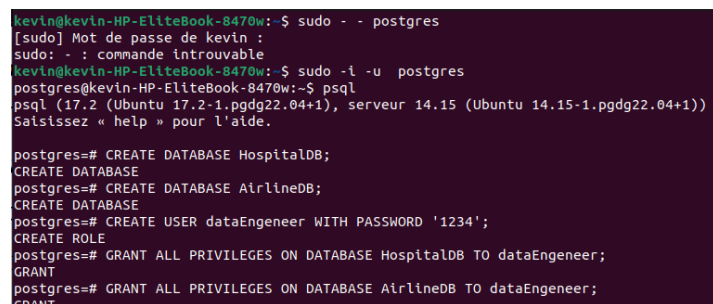
0.2.3 Création des bases de données

En supposant que vous avez PostgreSQL installé sur votre PC accédez à PostgreSQL via le terminal pour créer nos bases de données avec un propriétaire de la manière suivante:

- **CREATE DATABASE HospitalDB;**
- **CREATE DATABASE AirlineDB;**
- **CREATE USER dataEngineer WITH PASSWORD '1234';**

Attribuer tous les privilèges de ces bases de données au user dataEngineer :

- **GRANT ALL PRIVILEGES ON DATABASE AirlineDB TO dataEngineer;**
- **GRANT ALL PRIVILEGES ON DATABASE HospitalDB TO dataEngineer;**



```
kevin@kevin-HP-EliteBook-8470w:~$ sudo -i postgres
[sudo] Mot de passe de kevin :
sudo: -i : commande introuvable
kevin@kevin-HP-EliteBook-8470w:~$ sudo -i -u postgres
postgres@kevin-HP-EliteBook-8470w:~$ psql
psql (17.2 (Ubuntu 17.2-1.pgdg22.04+1), serveur 14.15 (Ubuntu 14.15-1.pgdg22.04+1))
Saisissez « help » pour l'aide.

postgres=# CREATE DATABASE HospitalDB;
CREATE DATABASE
postgres=# CREATE DATABASE AirlineDB;
CREATE DATABASE
postgres=# CREATE USER dataEngineer WITH PASSWORD '1234';
CREATE ROLE
postgres=# GRANT ALL PRIVILEGES ON DATABASE HospitalDB TO dataEngineer;
GRANT
postgres=# GRANT ALL PRIVILEGES ON DATABASE AirlineDB TO dataEngineer;
GRANT
```

Figure 1: capture d'écran des commandes

0.2. CONCEPTION DE LA BASE DE DONNÉES

Connection aux différentes bases de données sur pgAdmin pour mieux gérer l'espace de travail.

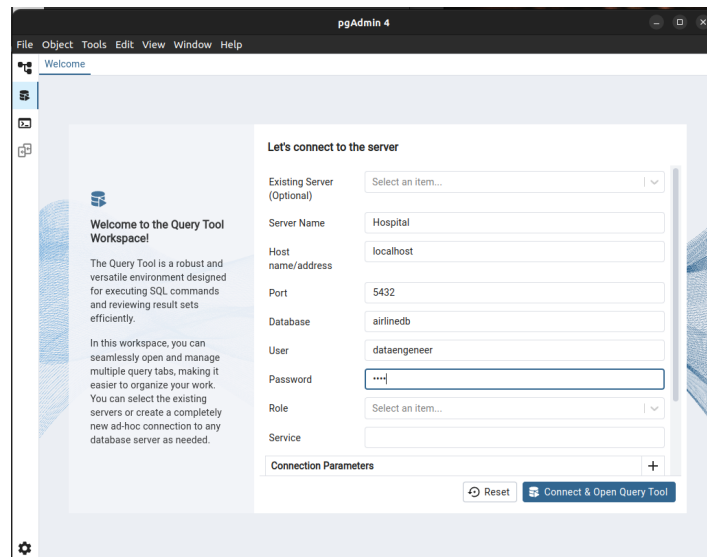


Figure 2: connexion a pgAdmin avec la BD AirlineDB

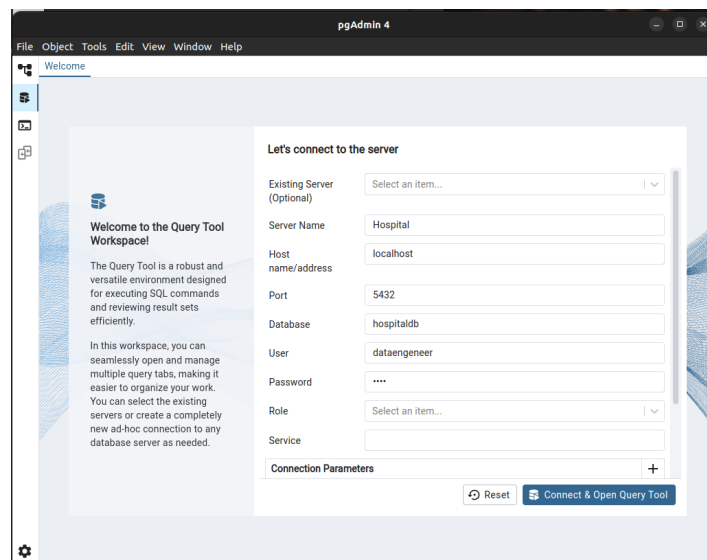


Figure 3: connexion a pgAdmin avec la BD HospitalDB

Remarque: PostgreSQL stocke les informations en lettre minuscule ainsi la base de données AirlineDB sera stockée comme [airlinedb](#), donc en remplissant les champs pour se connecter a pgAdmin il faudra en tenir compte.

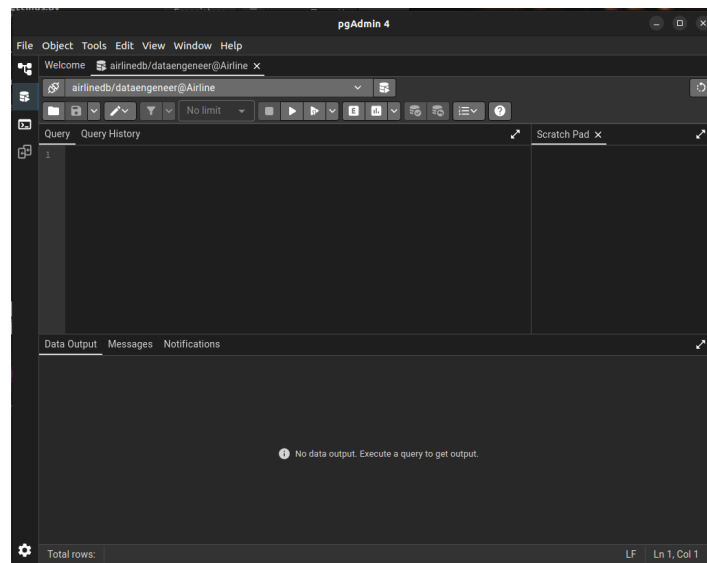


Figure 4: interface utilisateur de postgresQL

vous pouvez bien sur personnaliser l'interface selon vos besoins.

0.2.4 Creation des tables

Pour la Creation des tables on va proceder comme suite:

- selectionner la base de donnée [HospitalDB](#) et executez les requetes suivantes pour créer les tables:
 - ```
CREATE TABLE Patients (
 PatientID INT PRIMARY KEY,
 FirstName VARCHAR(50),
 LastName VARCHAR(50),
 DateOfBirth DATE,
 VisitDate DATE,
 Diagnosis VARCHAR(100)
);
```
  - ```
CREATE TABLE Doctors (  
  DoctorID INT PRIMARY KEY,  
  FirstName VARCHAR(50),  
  LastName VARCHAR(50),  
  Specialty VARCHAR(50)  
);
```
 - ```
CREATE TABLE Appointments (
 AppointmentID INT PRIMARY KEY,
 PatientID INT,
 DoctorID INT,
```

```
AppointmentDate DATE,
Notes VARCHAR(255),
FOREIGN KEY (PatientID) REFERENCES
Patients(PatientID),
FOREIGN KEY (DoctorID) REFERENCES
Doctors(DoctorID)
);
```

insertion de quelques données dans la base de donnée

```
INSERT INTO Patients (PatientID, FirstName, LastName, DateOfBirth, VisitDate, Di-
agnosis)
VALUES (1, 'Rahul', 'Sharma', '1985-04-23', '2024-08-01', 'Fever');
```

```
INSERT INTO Doctors (DoctorID, FirstName, LastName, Specialty)
VALUES (1, 'Sita', 'Verma', 'Cardiology');
```

```
INSERT INTO Appointments (AppointmentID, PatientID, DoctorID, AppointmentDate,
Notes)
VALUES (1, 1, 1, '2024-08-02', 'Routine check-up');
```

Mise a jour des données

```
UPDATE Patients
SET Diagnosis = 'Common Cold'
WHERE PatientID = 1;
```

```
UPDATE Doctors
SET Specialty = 'Pediatrics'
WHERE DoctorID = 1;
```

Suppression des données

```
DELETE FROM Patients
WHERE PatientID = 1;
```

```
DELETE FROM Appointments
WHERE AppointmentID = 1;
```

```
INSERT INTO Patients (PatientID, FirstName, LastName, DateOfBirth, VisitDate, Di-
agnosis)
VALUES
```



```
(1, 'Rahul', 'Sharma', '1985-04-23', '2024-08-01', 'Fever'),
(2, 'Priya', 'Kumar', '1990-07-11', '2024-08-02', 'Diabetes'),
(3, 'Amit', 'Singh', '1975-09-17', '2024-08-03', 'Hypertension');
```

```
INSERT INTO Doctors (DoctorID, FirstName, LastName, Specialty) VALUES
(1, 'Sita', 'Verma', 'Cardiology'),
(2, 'Arjun', 'Rao', 'Endocrinology'),
(3, 'Ravi', 'Gupta', 'General Practice');
```

```
INSERT INTO Appointments (AppointmentID, PatientID, DoctorID, AppointmentDate,
Notes) VALUES
(1, 1, 1, '2024-08-04', 'Follow-up visit for fever.'),
(2, 2, 2, '2024-08-05', 'Diabetes management check.'),
(3, 3, 3, '2024-08-06', 'Regular health check-up.');
```

ainsi je vous laisse le soin de remplir vos bases de données pour les travaux avenir.

- [selectionner la base de donnée AirlineDB](#) et exécutez les requetes suivantes pour créer les tables:

```
– CREATE TABLE Flights (
 FlightID INT PRIMARY KEY,
 FlightNumber VARCHAR(10),
 DepartureCity VARCHAR(50),
 ArrivalCity VARCHAR(50),
 DepartureDate DATE,
 Status VARCHAR(20)
);

– CREATE TABLE Passengers (
 PassengerID INT PRIMARY KEY,
 FirstName VARCHAR(50),
 LastName VARCHAR(50),
 DateOfBirth DATE
);

– CREATE TABLE Bookings (
 BookingID INT PRIMARY KEY,
 FlightID INT,
 PassengerID INT,
 BookingDate DATE,
 FOREIGN KEY (FlightID) REFERENCES Flights(FlightID),
 FOREIGN KEY (PassengerID) REFERENCES
 Passengers(PassengerID)
);
```

insertion des données dans la base de données:

```
INSERT INTO Flights (FlightID, FlightNumber, DepartureCity, ArrivalCity, De-
partureDate, Status)
VALUES
(1, 'AI202', 'Mumbai', 'Delhi', '2024-08-10', 'On Time'),
(2, 'AI203', 'Delhi', 'Bangalore', '2024-08-11', 'Delayed'),
(3, 'AI204', 'Chennai', 'Mumbai', '2024-08-12', 'On Time');
```

```
INSERT INTO Passengers (PassengerID, FirstName, LastName, DateOfBirth)
VALUES
(1, 'Anita', 'Desai', '1988-02-14'),
(2, 'Raj', 'Patel', '1992-03-22'),
(3, 'Simran', 'Kapoor', '1978-11-08');
```

```
INSERT INTO Bookings (BookingID, FlightID, PassengerID, BookingDate)
VALUES
(1, 1, 1, '2024-08-01'),
(2, 2, 2, '2024-08-02'),
(3, 3, 3, '2024-08-03');
```

### 0.3 quelques scénarios de la vie réelle pour un DataEngineer

#### 0.3.1 Trouver tous les patients avec un diagnostic spécifique

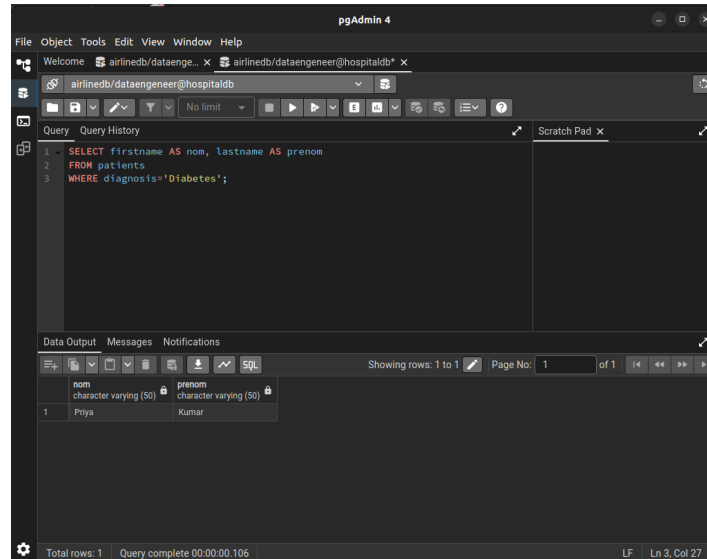


Figure 5: requete SQL

#### 0.3.2 Lister tous les rendez-vous pour un medecin spécifique

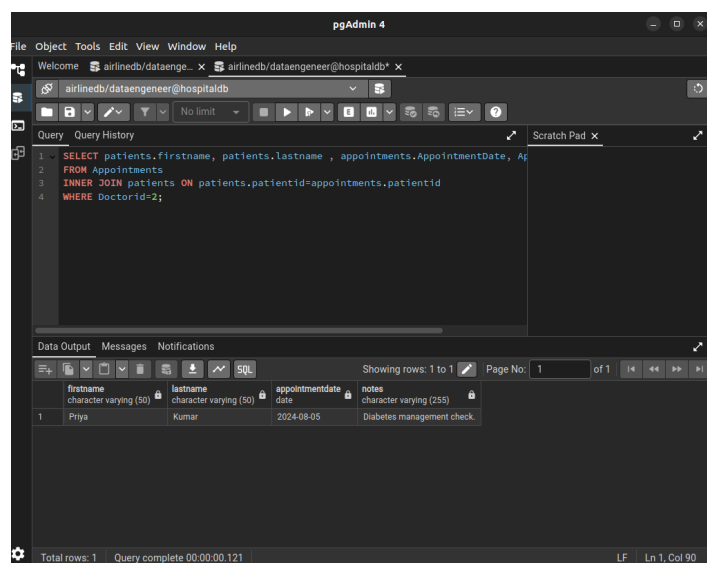


Figure 6: requete SQL



## 0.4 Sauvegarde Journalier des données de CamairCo et de l'hôpital central

### 0.4.1 Mise en situation

Supposons que les données relatives aux compagnies de CamerCo et ceux de l'hôpital central doivent être sauvegardées tous les jours à 12 heures 00 minutes.

### 0.4.2 Cas de CamairCo

Afin de répondre à ce problème nous allons créer une fonction python qui se chargera d'effectuer ce traitement tout seul de manière automatique.

Ainsi nous aurons besoin des bibliothèques suivantes:

- schedule
- Time
- pandas

Définissons cette fonction, on supposera que les données de CamairCo(AirlineDB) sont stockées dans un fichier csv

Le code jupyter se trouve dans le dossier Workshop

## 0.5 Collection des données

### 0.5.1 Collection des données a partir des fichiers

#### Types de fichiers

Aujourd'hui les données sont stockées dans différents formats de fichiers parmi lesquels

- fichiers CSV(Comma-Separated Values) avec l'extention .csv dans lequel les données sont séparées par des virgule. l'un des soucis avec ce format de fichier est qu'il ne peut pas prendre en charge les données avec des structures complexes.
- fichiers TXT(Plain text) avec l'extention .txt ce type de format nous permet de fixer un delimiteur de notre choix(dans le cas des fichiers csv, le délimiteur c'est le point virgule).
- fichiers XLSX(Microsoft Excel), c'est un format utilisé par microsoft pour stocker les données. il est généralement utilisé pour le "business analysis scenarios". cependant , excel presente uninconvenient qui est sa limite quant aux données que peut contenir un sheet, ce nombre de donné envoisine les 1050000 lignes de données.
- fichiers JSON (JavaScript Object Notation) , ce format peut contenir des données avec des structure complexe. et les données y sont sockées sous forme de dictionnaire: "clé:valeur".
- et bien d'autres fichier comme les fichier HTML pour les pages web, fichiers XML.

#### Cas pratiques

Cette section est une partie qui sera essentiellement pratiques. les données sont stockées dans le dossier datas.



Figure 8: fichier csv

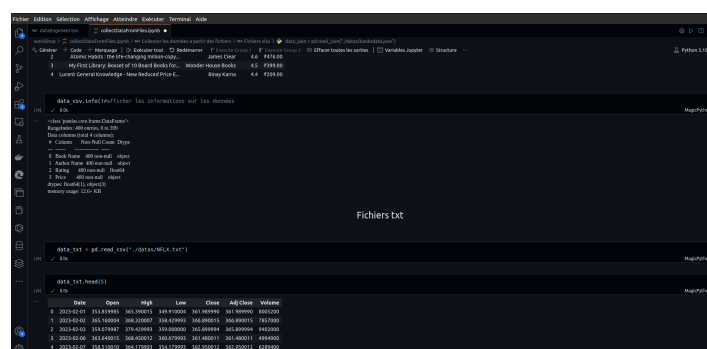
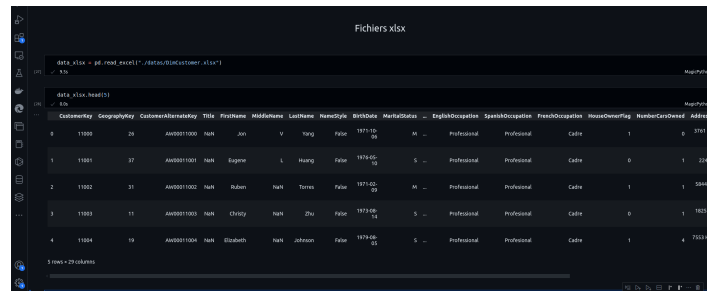


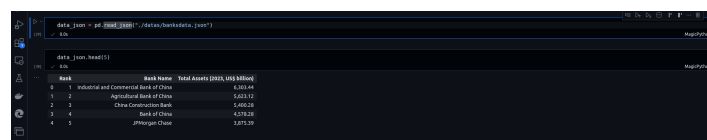
Figure 9: fichir txt



```
data_xlsx = pd.read_excel('data/customer.xlsx')
data_xlsx.head()
```

|   | CustomerKey | GeographyKey | CustomerAddressKey | Title | FirstName | MiddleName | LastName     | NameStyle | BirthDate  | MaritalStatus | EnglishOccupation | SpanishOccupation | FrenchOccupation | HouseOwnerFlag | NumberChildren | Address  |
|---|-------------|--------------|--------------------|-------|-----------|------------|--------------|-----------|------------|---------------|-------------------|-------------------|------------------|----------------|----------------|----------|
| 0 | 11000       | 28           | AMS00011000        | Ms    | Jean      |            | V. Wang      | False     | 1971-10-06 | M             | Professional      | Professional      | Cadre            | 1              | 0              | 2101 N.  |
| 1 | 11001       | 27           | AMS00011001        | Ms    | Eugene    |            | L. Huang     | False     | 1976-05-19 | S             | Professional      | Professional      | Cadre            | 0              | 1              | 2243 E.  |
| 2 | 11002       | 31           | AMS00011002        | Ms    | Ruben     |            | Mark Torres  | False     | 1977-02-01 | M             | Professional      | Professional      | Cadre            | 1              | 1              | 5864 E.  |
| 3 | 11003       | 11           | AMS00011003        | Ms    | Christy   |            | Mark Zhu     | False     | 1973-09-14 | S             | Professional      | Professional      | Cadre            | 0              | 1              | 1801 W.  |
| 4 | 11004       | 19           | AMS00011004        | Ms    | Elizabeth |            | Mark Johnson | False     | 1979-06-01 | S             | Professional      | Professional      | Cadre            | 1              | 4              | 7123 Ave |

Figure 10: fichier excel xlsx



```
data_json = pd.read_json('data/bankdata.json')
data_json.head()
```

|   | Bank | Bank Name                               | Total Assets (2021, US\$ Billions) |
|---|------|-----------------------------------------|------------------------------------|
| 0 | 1    | Industrial and Commercial Bank of China | 6,383.49                           |
| 1 | 2    | Agricultural Bank of China              | 5,653.19                           |
| 2 | 3    | China Construction Bank                 | 5,486.29                           |
| 3 | 4    | Bank of China                           | 5,079.26                           |
| 4 | 5    | JPMorgan Chase                          | 3,871.39                           |

Figure 11: fichier json

### 0.5.2 Collection des données a partir du web

#### Avantages de la collecte des données via le web

Le web contient une immense quantité de données de nature diverse. la collection des données via le web presente de nombreux avantages parmi lesquels

- [accès a une quantité d'information variée](#)
- [donées en temps réel et a jour](#)
- [intelligence compétitive](#) :en collectant les données a partir du web, on developpe des compétences et avons acces a des informations telles que suivi des activités en temps réel, les tendances du moment etc...
- [La recherche et l'analyse](#): La collection des données via le web est tes important pour les recherches, l'analyse des sentiments etc...

#### Structure des pages web

De nos jours , il existe plusieurs types de sites web tels que:

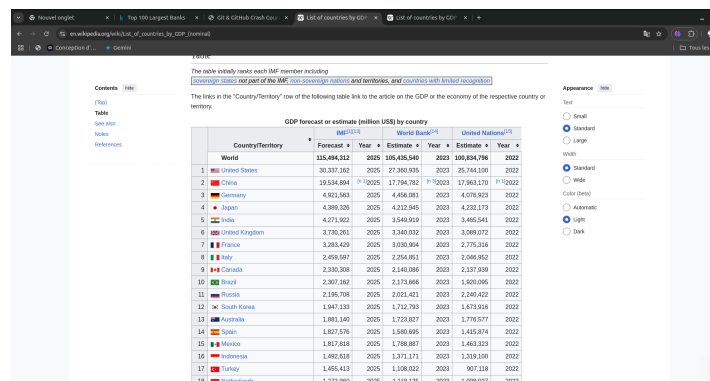
- [Les sites web structurés](#): ce sont des sites web organisés et ayant un format bien définis. ils sont repérés par une mise en page de bonne qualité, des sections clairement definies pour extraire les données on pourra utiliser les librairies comme BeautifulSoup ou lxml de python.
- [Les sites web semi-structurés](#): ils contiennent les données structurées et non structurées. Dans ce genre de site web, certaines section sont bien structurées et d'autre ne le sont pas.

- **les sites web non structurés** : Ils n'ont pas une organisation claire ou une structure bien définie. Ce genre de site constitue un vrai challenge car les données n'y sont pas structurées et sont de nature diverse telles que les images, les formats de données non tabulaires etc... Tout cela requiert donc les techniques sophistiquées pour leur extraction. Spécialement pour ce genre de sites on peut utiliser les techniques de traitement du langage naturel (NLP) ou les algorithmes de machine learning

### 0.5.3 Cas pratiques: scraper les données météo

#### Sites web ayant une structure facile à lire

considérons le site web wikipedia '[https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_GDP\\_\(nominal\)#Table](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)#Table)' qui contient les données sur les PIB des pays et essayons d'y extraire les données. Le '#table' permet de pointer vers un élément spécifique de la page web.



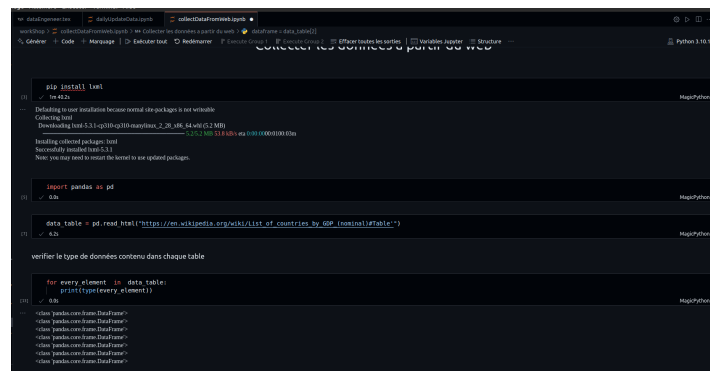
The table initially nests each IMF member including (coverage listed as part of the IMF, not coverage of nations and territories, and countries with limited recognition)

The links in the "Country/Territory" row of the following table link to the article on the GDP or the economy of the respective country or territory.

|    |                                                           | GDP forecast or estimate (billion US\$) by country |      |                   |      |
|----|-----------------------------------------------------------|----------------------------------------------------|------|-------------------|------|
|    |                                                           | IMF (2021)                                         |      | World Bank (2021) |      |
|    |                                                           | Forecast                                           | Year | Estimate          | Year |
| 1  | <span><span></span></span> <a href="#">World</a>          | 115,494,312                                        | 2025 | 105,435,540       | 2023 |
| 2  | <span><span></span></span> <a href="#">United States</a>  | 20,247,192                                         | 2025 | 27,360,035        | 2023 |
| 3  | <span><span></span></span> <a href="#">China</a>          | 10,524,604                                         | 2025 | 17,704,752        | 2023 |
| 4  | <span><span></span></span> <a href="#">Germany</a>        | 4,021,563                                          | 2025 | 4,456,081         | 2023 |
| 5  | <span><span></span></span> <a href="#">India</a>          | 4,389,326                                          | 2025 | 4,232,945         | 2023 |
| 6  | <span><span></span></span> <a href="#">United Kingdom</a> | 4,271,922                                          | 2025 | 3,549,919         | 2023 |
| 7  | <span><span></span></span> <a href="#">France</a>         | 3,720,261                                          | 2025 | 3,340,032         | 2023 |
| 8  | <span><span></span></span> <a href="#">Italy</a>          | 3,285,429                                          | 2025 | 3,030,904         | 2023 |
| 9  | <span><span></span></span> <a href="#">Canada</a>         | 2,459,597                                          | 2025 | 2,254,051         | 2023 |
| 10 | <span><span></span></span> <a href="#">Brazil</a>         | 2,330,308                                          | 2025 | 2,140,086         | 2023 |
| 11 | <span><span></span></span> <a href="#">Russia</a>         | 2,307,192                                          | 2025 | 2,175,086         | 2023 |
| 12 | <span><span></span></span> <a href="#">South Korea</a>    | 2,185,708                                          | 2025 | 2,021,451         | 2023 |
| 13 | <span><span></span></span> <a href="#">Australia</a>      | 1,947,133                                          | 2025 | 1,712,783         | 2023 |
| 14 | <span><span></span></span> <a href="#">Spain</a>          | 1,901,140                                          | 2025 | 1,723,027         | 2023 |
| 15 | <span><span></span></span> <a href="#">Mexico</a>         | 1,827,570                                          | 2025 | 1,580,095         | 2023 |
| 16 | <span><span></span></span> <a href="#">Indonesia</a>      | 1,817,618                                          | 2025 | 1,708,987         | 2023 |
| 17 | <span><span></span></span> <a href="#">Turkey</a>         | 1,465,413                                          | 2025 | 1,371,171         | 2023 |
| 18 | <span><span></span></span> <a href="#">Netherlands</a>    | 1,455,413                                          | 2025 | 1,108,032         | 2023 |

Figure 12: Contenu du site web où on fera l'extraction des données

les photos ci dessous illustrent le processus d'extraction de ces données



```

1 pip install bs4
2
3 # Importing libraries
4 from bs4 import BeautifulSoup
5 import pandas as pd
6
7 # Reading the HTML content from the URL
8 data_table = pd.read_html("https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)#Table")
9
10 # Verifying the type of data contained in each table
11 for every_element in data_table:
12 print(type(every_element))
13
14 # Displaying the first table
15 data_table[0]
16
17 # Displaying the second table
18 data_table[1]
19
20 # Displaying the third table
21 data_table[2]
22
23 # Displaying the fourth table
24 data_table[3]
25
26 # Displaying the fifth table
27 data_table[4]
28
29 # Displaying the sixth table
30 data_table[5]
31
32 # Displaying the seventh table
33 data_table[6]
34
35 # Displaying the eighth table
36 data_table[7]
37
38 # Displaying the ninth table
39 data_table[8]
40
41 # Displaying the tenth table
42 data_table[9]
43
44 # Displaying the eleventh table
45 data_table[10]
46
47 # Displaying the twelfth table
48 data_table[11]
49
50 # Displaying the thirteenth table
51 data_table[12]
52
53 # Displaying the fourteenth table
54 data_table[13]
55
56 # Displaying the fifteenth table
57 data_table[14]
58
59 # Displaying the sixteenth table
60 data_table[15]
61
62 # Displaying the seventeenth table
63 data_table[16]
64
65 # Displaying the eighteenth table
66 data_table[17]
67
68 # Displaying the nineteenth table
69 data_table[18]
70
71 # Displaying the twentieth table
72 data_table[19]
73
74 # Displaying the twenty-first table
75 data_table[20]
76
77 # Displaying the twenty-second table
78 data_table[21]
79
80 # Displaying the twenty-third table
81 data_table[22]
82
83 # Displaying the twenty-fourth table
84 data_table[23]
85
86 # Displaying the twenty-fifth table
87 data_table[24]
88
89 # Displaying the twenty-sixth table
90 data_table[25]
91
92 # Displaying the twenty-seventh table
93 data_table[26]
94
95 # Displaying the twenty-eighth table
96 data_table[27]
97
98 # Displaying the twenty-ninth table
99 data_table[28]
100
101 # Displaying the thirtieth table
102 data_table[29]
103
104 # Displaying the thirty-first table
105 data_table[30]
106
107 # Displaying the thirty-second table
108 data_table[31]
109
110 # Displaying the thirty-third table
111 data_table[32]
112
113 # Displaying the thirty-fourth table
114 data_table[33]
115
116 # Displaying the thirty-fifth table
117 data_table[34]
118
119 # Displaying the thirty-sixth table
120 data_table[35]
121
122 # Displaying the thirty-seventh table
123 data_table[36]
124
125 # Displaying the thirty-eighth table
126 data_table[37]
127
128 # Displaying the thirty-ninth table
129 data_table[38]
130
131 # Displaying the fortieth table
132 data_table[39]
133
134 # Displaying the forty-first table
135 data_table[40]
136
137 # Displaying the forty-second table
138 data_table[41]
139
140 # Displaying the forty-third table
141 data_table[42]
142
143 # Displaying the forty-fourth table
144 data_table[43]
145
146 # Displaying the forty-fifth table
147 data_table[44]
148
149 # Displaying the forty-sixth table
150 data_table[45]
151
152 # Displaying the forty-seventh table
153 data_table[46]
154
155 # Displaying the forty-eighth table
156 data_table[47]
157
158 # Displaying the forty-ninth table
159 data_table[48]
160
161 # Displaying the fiftieth table
162 data_table[49]
163
164 # Displaying the fifty-first table
165 data_table[50]
166
167 # Displaying the fifty-second table
168 data_table[51]
169
170 # Displaying the fifty-third table
171 data_table[52]
172
173 # Displaying the fifty-fourth table
174 data_table[53]
175
176 # Displaying the fifty-fifth table
177 data_table[54]
178
179 # Displaying the fifty-sixth table
180 data_table[55]
181
182 # Displaying the fifty-seventh table
183 data_table[56]
184
185 # Displaying the fifty-eighth table
186 data_table[57]
187
188 # Displaying the fifty-ninth table
189 data_table[58]
190
191 # Displaying the sixtieth table
192 data_table[59]
193
194 # Displaying the sixty-first table
195 data_table[60]
196
197 # Displaying the sixty-second table
198 data_table[61]
199
200 # Displaying the sixty-third table
201 data_table[62]
202
203 # Displaying the sixty-fourth table
204 data_table[63]
205
206 # Displaying the sixty-fifth table
207 data_table[64]
208
209 # Displaying the sixty-sixth table
210 data_table[65]
211
212 # Displaying the sixty-seventh table
213 data_table[66]
214
215 # Displaying the sixty-eighth table
216 data_table[67]
217
218 # Displaying the sixty-ninth table
219 data_table[68]
220
221 # Displaying the seventieth table
222 data_table[69]
223
224 # Displaying the seventy-first table
225 data_table[70]
226
227 # Displaying the seventy-second table
228 data_table[71]
229
230 # Displaying the seventy-third table
231 data_table[72]
232
233 # Displaying the seventy-fourth table
234 data_table[73]
235
236 # Displaying the seventy-fifth table
237 data_table[74]
238
239 # Displaying the seventy-sixth table
240 data_table[75]
241
242 # Displaying the seventy-seventh table
243 data_table[76]
244
245 # Displaying the seventy-eighth table
246 data_table[77]
247
248 # Displaying the seventy-ninth table
249 data_table[78]
250
251 # Displaying the eightieth table
252 data_table[79]
253
254 # Displaying the eighty-first table
255 data_table[80]
256
257 # Displaying the eighty-second table
258 data_table[81]
259
260 # Displaying the eighty-third table
261 data_table[82]
262
263 # Displaying the eighty-fourth table
264 data_table[83]
265
266 # Displaying the eighty-fifth table
267 data_table[84]
268
269 # Displaying the eighty-sixth table
270 data_table[85]
271
272 # Displaying the eighty-seventh table
273 data_table[86]
274
275 # Displaying the eighty-eighth table
276 data_table[87]
277
278 # Displaying the eighty-ninth table
279 data_table[88]
280
281 # Displaying the ninetieth table
282 data_table[89]
283
284 # Displaying the ninety-first table
285 data_table[90]
286
287 # Displaying the ninety-second table
288 data_table[91]
289
290 # Displaying the ninety-third table
291 data_table[92]
292
293 # Displaying the ninety-fourth table
294 data_table[93]
295
296 # Displaying the ninety-fifth table
297 data_table[94]
298
299 # Displaying the ninety-sixth table
300 data_table[95]
301
302 # Displaying the ninety-seventh table
303 data_table[96]
304
305 # Displaying the ninety-eighth table
306 data_table[97]
307
308 # Displaying the ninety-ninth table
309 data_table[98]
310
311 # Displaying the hundredth table
312 data_table[99]
313
314 # Displaying the hundred-first table
315 data_table[100]
316
317 # Displaying the hundred-second table
318 data_table[101]
319
320 # Displaying the hundred-third table
321 data_table[102]
322
323 # Displaying the hundred-fourth table
324 data_table[103]
325
326 # Displaying the hundred-fifth table
327 data_table[104]
328
329 # Displaying the hundred-sixth table
330 data_table[105]
331
332 # Displaying the hundred-seventh table
333 data_table[106]
334
335 # Displaying the hundred-eighth table
336 data_table[107]
337
338 # Displaying the hundred-ninth table
339 data_table[108]
340
341 # Displaying the hundred-tenth table
342 data_table[109]
343
344 # Displaying the hundred-eleventh table
345 data_table[110]
346
347 # Displaying the hundred-twelfth table
348 data_table[111]
349
350 # Displaying the hundred-thirteenth table
351 data_table[112]
352
353 # Displaying the hundred-fourteenth table
354 data_table[113]
355
356 # Displaying the hundred-fifteenth table
357 data_table[114]
358
359 # Displaying the hundred-sixteenth table
360 data_table[115]
361
362 # Displaying the hundred-seventeenth table
363 data_table[116]
364
365 # Displaying the hundred-eighteenth table
366 data_table[117]
367
368 # Displaying the hundred-nineteenth table
369 data_table[118]
370
371 # Displaying the hundred-twentieth table
372 data_table[119]
373
374 # Displaying the hundred-twenty-first table
375 data_table[120]
376
377 # Displaying the hundred-twenty-second table
378 data_table[121]
379
380 # Displaying the hundred-twenty-third table
381 data_table[122]
382
383 # Displaying the hundred-twenty-fourth table
384 data_table[123]
385
386 # Displaying the hundred-twenty-fifth table
387 data_table[124]
388
389 # Displaying the hundred-twenty-sixth table
389 data_table[125]
390
391 # Displaying the hundred-twenty-seventh table
391 data_table[126]
392
393 # Displaying the hundred-twenty-eighth table
393 data_table[127]
394
395 # Displaying the hundred-twenty-ninth table
395 data_table[128]
396
397 # Displaying the hundred-thirtieth table
397 data_table[129]
398
399 # Displaying the hundred-thirty-first table
399 data_table[130]
400
401 # Displaying the hundred-thirty-second table
401 data_table[131]
402
403 # Displaying the hundred-thirty-third table
403 data_table[132]
404
405 # Displaying the hundred-thirty-fourth table
405 data_table[133]
406
407 # Displaying the hundred-thirty-fifth table
407 data_table[134]
408
409 # Displaying the hundred-thirty-sixth table
409 data_table[135]
410
411 # Displaying the hundred-thirty-seventh table
411 data_table[136]
412
413 # Displaying the hundred-thirty-eighth table
413 data_table[137]
414
415 # Displaying the hundred-thirty-ninth table
415 data_table[138]
416
417 # Displaying the hundred-fortieth table
417 data_table[139]
418
419 # Displaying the hundred-forty-first table
419 data_table[140]
420
421 # Displaying the hundred-forty-second table
421 data_table[141]
422
423 # Displaying the hundred-forty-third table
423 data_table[142]
424
425 # Displaying the hundred-forty-fourth table
425 data_table[143]
426
427 # Displaying the hundred-forty-fifth table
427 data_table[144]
428
429 # Displaying the hundred-forty-sixth table
429 data_table[145]
430
431 # Displaying the hundred-forty-seventh table
431 data_table[146]
432
433 # Displaying the hundred-forty-eighth table
433 data_table[147]
434
435 # Displaying the hundred-forty-ninth table
435 data_table[148]
436
437 # Displaying the hundred-fiftieth table
437 data_table[149]
438
439 # Displaying the hundred-fifty-first table
439 data_table[150]
440
441 # Displaying the hundred-fifty-second table
441 data_table[151]
442
443 # Displaying the hundred-fifty-third table
443 data_table[152]
444
445 # Displaying the hundred-fifty-fourth table
445 data_table[153]
446
447 # Displaying the hundred-fifty-fifth table
447 data_table[154]
448
449 # Displaying the hundred-fifty-sixth table
449 data_table[155]
450
451 # Displaying the hundred-fifty-seventh table
451 data_table[156]
452
453 # Displaying the hundred-fifty-eighth table
453 data_table[157]
454
455 # Displaying the hundred-fifty-ninth table
455 data_table[158]
456
457 # Displaying the hundred-sixtieth table
457 data_table[159]
458
459 # Displaying the hundred-sixty-first table
459 data_table[160]
460
461 # Displaying the hundred-sixty-second table
461 data_table[161]
462
463 # Displaying the hundred-sixty-third table
463 data_table[162]
464
465 # Displaying the hundred-sixty-fourth table
465 data_table[163]
466
467 # Displaying the hundred-sixty-fifth table
467 data_table[164]
468
469 # Displaying the hundred-sixty-sixth table
469 data_table[165]
470
471 # Displaying the hundred-sixty-seventh table
471 data_table[166]
472
473 # Displaying the hundred-sixty-eighth table
473 data_table[167]
474
475 # Displaying the hundred-sixty-ninth table
475 data_table[168]
476
477 # Displaying the hundred-seventieth table
477 data_table[169]
478
479 # Displaying the hundred-seventy-first table
479 data_table[170]
480
481 # Displaying the hundred-seventy-second table
481 data_table[171]
482
483 # Displaying the hundred-seventy-third table
483 data_table[172]
484
485 # Displaying the hundred-seventy-fourth table
485 data_table[173]
486
487 # Displaying the hundred-seventy-fifth table
487 data_table[174]
488
489 # Displaying the hundred-seventy-sixth table
489 data_table[175]
490
491 # Displaying the hundred-seventy-seventh table
491 data_table[176]
492
493 # Displaying the hundred-seventy-eighth table
493 data_table[177]
494
495 # Displaying the hundred-seventy-ninth table
495 data_table[178]
496
497 # Displaying the hundred-eightieth table
497 data_table[179]
498
499 # Displaying the hundred-eighty-first table
499 data_table[180]
500
501 # Displaying the hundred-eighty-second table
501 data_table[181]
502
503 # Displaying the hundred-eighty-third table
503 data_table[182]
504
505 # Displaying the hundred-eighty-fourth table
505 data_table[183]
506
507 # Displaying the hundred-eighty-fifth table
507 data_table[184]
508
509 # Displaying the hundred-eighty-sixth table
509 data_table[185]
510
511 # Displaying the hundred-eighty-seventh table
511 data_table[186]
512
513 # Displaying the hundred-eighty-eighth table
513 data_table[187]
514
515 # Displaying the hundred-eighty-ninth table
515 data_table[188]
516
517 # Displaying the hundred-ninetieth table
517 data_table[189]
518
519 # Displaying the hundred-ninety-first table
519 data_table[190]
520
521 # Displaying the hundred-ninety-second table
521 data_table[191]
522
523 # Displaying the hundred-ninety-third table
523 data_table[192]
524
525 # Displaying the hundred-ninety-fourth table
525 data_table[193]
526
527 # Displaying the hundred-ninety-fifth table
527 data_table[194]
528
529 # Displaying the hundred-ninety-sixth table
529 data_table[195]
530
531 # Displaying the hundred-ninety-seventh table
531 data_table[196]
532
533 # Displaying the hundred-ninety-eighth table
533 data_table[197]
534
535 # Displaying the hundred-ninety-ninth table
535 data_table[198]
536
537 # Displaying the hundredth table
537 data_table[199]
538
539 # Displaying the hundred-first table
539 data_table[200]
540
541 # Displaying the hundred-second table
541 data_table[201]
542
543 # Displaying the hundred-third table
543 data_table[202]
544
545 # Displaying the hundred-fourth table
545 data_table[203]
546
547 # Displaying the hundred-fifth table
547 data_table[204]
548
549 # Displaying the hundred-sixth table
549 data_table[205]
550
551 # Displaying the hundred-seventh table
551 data_table[206]
552
553 # Displaying the hundred-eighth table
553 data_table[207]
554
555 # Displaying the hundred-ninth table
555 data_table[208]
556
557 # Displaying the hundred-tenth table
557 data_table[209]
558
559 # Displaying the hundred-eleventh table
559 data_table[210]
560
561 # Displaying the hundred-twelfth table
561 data_table[211]
562
563 # Displaying the hundred-thirteenth table
563 data_table[212]
564
565 # Displaying the hundred-fourteenth table
565 data_table[213]
566
567 # Displaying the hundred-fifteenth table
567 data_table[214]
568
569 # Displaying the hundred-sixteenth table
569 data_table[215]
570
571 # Displaying the hundred-seventeenth table
571 data_table[216]
572
573 # Displaying the hundred-eighteenth table
573 data_table[217]
574
575 # Displaying the hundred-nineteenth table
575 data_table[218]
576
577 # Displaying the hundred-twentieth table
577 data_table[219]
578
579 # Displaying the hundred-twenty-first table
579 data_table[220]
580
581 # Displaying the hundred-twenty-second table
581 data_table[221]
582
583 # Displaying the hundred-twenty-third table
583 data_table[222]
584
585 # Displaying the hundred-twenty-fourth table
585 data_table[223]
586
587 # Displaying the hundred-twenty-fifth table
587 data_table[224]
588
589 # Displaying the hundred-twenty-sixth table
589 data_table[225]
590
591 # Displaying the hundred-twenty-seventh table
591 data_table[226]
592
593 # Displaying the hundred-twenty-eighth table
593 data_table[227]
594
595 # Displaying the hundred-twenty-ninth table
595 data_table[228]
596
597 # Displaying the hundred-thirtieth table
597 data_table[229]
598
599 # Displaying the hundred-thirty-first table
599 data_table[230]
600
601 # Displaying the hundred-thirty-second table
601 data_table[231]
602
603 # Displaying the hundred-thirty-third table
603 data_table[232]
604
605 # Displaying the hundred-thirty-fourth table
605 data_table[233]
606
607 # Displaying the hundred-thirty-fifth table
607 data_table[234]
608
609 # Displaying the hundred-thirty-sixth table
609 data_table[235]
610
611 # Displaying the hundred-thirty-seventh table
611 data_table[236]
612
613 # Displaying the hundred-thirty-eighth table
613 data_table[237]
614
615 # Displaying the hundred-thirty-ninth table
615 data_table[238]
616
617 # Displaying the hundred-fortieth table
617 data_table[239]
618
619 # Displaying the hundred-forty-first table
619 data_table[240]
620
621 # Displaying the hundred-forty-second table
621 data_table[241]
622
623 # Displaying the hundred-forty-third table
623 data_table[242]
624
625 # Displaying the hundred-forty-fourth table
625 data_table[243]
626
627 # Displaying the hundred-forty-fifth table
627 data_table[244]
628
629 # Displaying the hundred-forty-sixth table
629 data_table[245]
630
631 # Displaying the hundred-forty-seventh table
631 data_table[246]
632
633 # Displaying the hundred-forty-eighth table
633 data_table[247]
634
635 # Displaying the hundred-forty-ninth table
635 data_table[248]
636
637 # Displaying the hundred-fiftieth table
637 data_table[249]
638
639 # Displaying the hundred-fifty-first table
639 data_table[250]
640
641 # Displaying the hundred-fifty-second table
641 data_table[251]
642
643 # Displaying the hundred-fifty-third table
643 data_table[252]
644
645 # Displaying the hundred-fifty-fourth table
645 data_table[253]
646
647 # Displaying the hundred-fifty-fifth table
647 data_table[254]
648
649 # Displaying the hundred-fifty-sixth table
649 data_table[255]
650
651 # Displaying the hundred-fifty-seventh table
651 data_table[256]
652
653 # Displaying the hundred-fifty-eighth table
653 data_table[257]
654
655 # Displaying the hundred-fifty-ninth table
655 data_table[258]
656
657 # Displaying the hundred-sixtieth table
657 data_table[259]
658
659 # Displaying the hundred-sixty-first table
659 data_table[260]
660
661 # Displaying the hundred-sixty-second table
661 data_table[261]
662
663 # Displaying the hundred-sixty-third table
663 data_table[262]
664
665 # Displaying the hundred-sixty-fourth table
665 data_table[263]
666
667 # Displaying the hundred-sixty-fifth table
667 data_table[264]
668
669 # Displaying the hundred-sixty-sixth table
669 data_table[265]
670
671 # Displaying the hundred-sixty-seventh table
671 data_table[266]
672
673 # Displaying the hundred-sixty-eighth table
673 data_table[267]
674
675 # Displaying the hundred-sixty-ninth table
675 data_table[268]
676
677 # Displaying the hundred-seventieth table
677 data_table[269]
678
679 # Displaying the hundred-seventy-first table
679 data_table[270]
680
681 # Displaying the hundred-seventy-second table
681 data_table[271]
682
683 # Displaying the hundred-seventy-third table
683 data_table[272]
684
685 # Displaying the hundred-seventy-fourth table
685 data_table[273]
686
687 # Displaying the hundred-seventy-fifth table
687 data_table[274]
688
689 # Displaying the hundred-seventy-sixth table
689 data_table[275]
690
691 # Displaying the hundred-seventy-seventh table
691 data_table[276]
692
693 # Displaying the hundred-seventy-eighth table
693 data_table[277]
694
695 # Displaying the hundred-seventy-ninth table
695 data_table[278]
696
697 # Displaying the hundred-eightieth table
697 data_table[279]
698
699 # Displaying the hundred-eighty-first table
699 data_table[280]
700
701 # Displaying the hundred-eighty-second table
701 data_table[281]
702
703 # Displaying the hundred-eighty-third table
703 data_table[282]
704
705 # Displaying the hundred-eighty-fourth table
705 data_table[283]
706
707 # Displaying the hundred-eighty-fifth table
707 data_table[284]
708
709 # Displaying the hundred-eighty-sixth table
709 data_table[285]
710
711 # Displaying the hundred-eighty-seventh table
711 data_table[286]
712
713 # Displaying the hundred-eighty-eighth table
713 data_table[287]
714
715 # Displaying the hundred-eighty-ninth table
715 data_table[288]
716
717 # Displaying the hundred-ninetieth table
717 data_table[289]
718
719 # Displaying the hundred-ninety-first table
719 data_table[290]
720
721 # Displaying the hundred-ninety-second table
721 data_table[291]
722
723 # Displaying the hundred-ninety-third table
723 data_table[292]
724
725 # Displaying the hundred-ninety-fourth table
725 data_table[293]
726
727 # Displaying the hundred-ninety-fifth table
727 data_table[294]
728
729 # Displaying the hundred-ninety-sixth table
729 data_table[295]
730
731 # Displaying the hundred-ninety-seventh table
731 data_table[296]
732
733 # Displaying the hundred-ninety-eighth table
733 data_table[297]
734
735 # Displaying the hundred-ninety-ninth table
735 data_table[298]
736
737 # Displaying the hundredth table
737 data_table[299]
738
739 # Displaying the hundred-first table
739 data_table[300]
740
741 # Displaying the hundred-second table
741 data_table[301]
742
743 # Displaying the hundred-third table
743 data_table[302]
744
745 # Displaying the hundred-fourth table
745 data_table[303]
746
747 # Displaying the hundred-fifth table
747 data_table[304]
748
749 # Displaying the hundred-sixth table
749 data_table[305]
750
751 # Displaying the hundred-seventh table
751 data_table[306]
752
753 # Displaying the hundred-eighth table
753 data_table[307]
754
755 # Displaying the hundred-ninth table
755 data_table[308]
756
757 # Displaying the hundred-tenth table
757 data_table[309]
758
759 # Displaying the hundred-eleventh table
759 data_table[310]
760
761 # Displaying the hundred-twelfth table
761 data_table[311]
762
763 # Displaying the hundred-thirteenth table
763 data_table[312]
764
765 # Displaying the hundred-fourteenth table
765 data_table[313]
766
767 # Displaying the hundred-fifteenth table
767 data_table[314]
768
769 # Displaying the hundred-sixteenth table
769 data_table[315]
770
771 # Displaying the hundred-seventeenth table
771 data_table[316]
772
773 # Displaying the hundred-eighteenth table
773 data_table[317]
774
775 # Displaying the hundred-nineteenth table
775 data_table[318]
776
777 # Displaying the hundred-twentieth table
777 data_table[319]
778
779 # Displaying the hundred-twenty-first table
779 data_table[320]
780
781 # Displaying the hundred-twenty-second table
781 data_table[321]
782
783 # Displaying the hundred-twenty-third table
783 data_table[322]
784
785 # Displaying the hundred-twenty-fourth table
785 data_table[323]
786
787 # Displaying the hundred-twenty-fifth table
787 data_table[324]
788
789 # Displaying the hundred-twenty-sixth table
789 data_table[325]
790
791 # Displaying the hundred-twenty-seventh table
791 data_table[326]
792
793 # Displaying the hundred-twenty-eighth table
793 data_table[327]
794
795 # Displaying the hundred-twenty-ninth table
795 data_table[328]
796
797 # Displaying the hundred-thirtieth table
797 data_table[329]
798
799 # Displaying the hundred-thirty-first table
799 data_table[330]
800
801 # Displaying the hundred-thirty-second table
801 data_table[331]
802
803 # Displaying the hundred-thirty-third table
803 data_table[332]
804
805 # Displaying the hundred-thirty-fourth table
805 data_table[333]
806
807 # Displaying the hundred-thirty-fifth table
807 data_table[334]
808
809 # Displaying the hundred-thirty-sixth table
809 data_table[335]
810
811 # Displaying the hundred-thirty-seventh table
811 data_table[336]
812
813 # Displaying the hundred-thirty-eighth table
813 data_table[337]
814
815 # Displaying the hundred-thirty-ninth table
815 data_table[338]
816
817 # Displaying the hundred-fortieth table
817 data_table[339]
818
819 # Displaying the hundred-forty-first table
819 data_table[340]
820
821 # Displaying the hundred-forty-second table
821 data_table[341]
822
823 # Displaying the hundred-forty-third table
823 data_table[342]
824
825 # Displaying the hundred-forty-fourth table
825 data_table[343]
826
827 # Displaying the hundred-forty-fifth table
827 data_table[344]
828
829 # Displaying the hundred-forty-sixth table
829 data_table[345]
830
831 # Displaying the hundred-forty-seventh table
831 data_table[346]
832
833 # Displaying the hundred-forty-eighth table
833 data_table[347]
834
835 # Displaying the hundred-forty-ninth table
835 data_table[348]
836
837 # Displaying the hundred-fiftieth table
837 data_table[349]
838
839 # Displaying the hundred-fifty-first table
839 data_table[350]
840
841 # Displaying the hundred-fifty-second table
841 data_table[351]
842
843 # Displaying the hundred-fifty-third table
843 data_table[352]
844
845 # Displaying the hundred-fifty-fourth table
845 data_table[353]
846
847 # Displaying the hundred-fifty-fifth table
847 data_table[354]
848
849 # Displaying the hundred-fifty-sixth table
849 data_table[355]
850
851 # Displaying the hundred-fifty-seventh table
851 data_table[356]
852
853 # Displaying the hundred-fifty-eighth table
853 data_table[357]
854
855 # Displaying the hundred-fifty-ninth table
855 data_table[358]
856
857 # Displaying the hundred-sixtieth table
857 data_table[359]
858
859 # Displaying the hundred-sixty-first table
859 data_table[360]
860
861 # Displaying the hundred-sixty-second table
861 data_table[361]
862
863 # Displaying the hundred-sixty-third table
863 data_table[362]
864
865 # Displaying the hundred-sixty-fourth table
865 data_table[363]
8
```



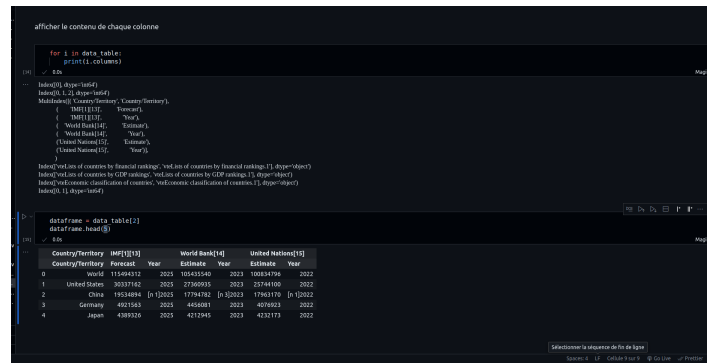


Figure 14: collection des données a partir du web

## Web Scraping : SCRAPPER LES DONNÉES METÉOS

certaines pages web sont semi-structurées contenant des méta-données comme les classes, labels et bien d'autres. une des méthodes pour y extraire les données est de visualiser le code source de la page web et de faire le web scraping. les compétences basiques en HTML et XML sont nécessaires pour cette partie. nous aurons donc besoin de télécharger la page web, et pour cela il existe plusieurs types de requêtes pour le faire : GET, POST etc...

on va télécharger la page

'<https://forecast.weather.gov/MapClick.php?lat=40.0466&lon=-105.2523#.YwpRBy2B1f0>'

qui est un site web de prévisions météo

En utilisant la méthode GET de requests de python.

La raison pour laquelle on ne va pas utiliser le site de wikipedia précédent est que wikipedia limite le web scraping et charge la plus part de ses pages avec ajax et de manière dynamique. Bibliothèques nécessaires:

- br4 de python
- requests

le scraping étant un exercice pratique, je vous invite à accéder au dossier workshop pour suivre le tutoriel via le fichier notebook en accédant à ce dossier via mon compte github énoncé en première page.

### 0.5.4 Collection des données à partir des bases de données

Pour le faire on va considérer nos bases de données [AirlineDB](#) et [HospitalDB](#) créées sous PostgreSQL au début de ce tutoriel ainsi que les identifiants du user des bases de données. ci-dessous est indiqué la méthode pour se connecter à postgresql sous python.

```
import psycopg2
conn = psycopg2.connect(database="airlinedb",
 user="dataengineer",
 host="localhost",
 password="1234",
 port="5432")
```

La suite de cette partie se fera de manière pratique, je vous invite à avoir accès au fichier notebook [collectDataFromDatabase.ipynb](#)

### 0.5.5 Collection des données en utilisant les API

La collection des données depuis les API (Application Programming Interface) présente de nombreux avantages parmi lesquels:

- Les données structurées : les API fournissent une structure standard de formats rendant ainsi facile l'établissement d'un pipeline de données.
- Données mises à jour en temps réel : elles fournissent les informations en temps réel permettant d'avoir accès aux dernières informations à jour.
- Contrôle d'accès : elles nous permettent d'avoir accès à leurs données en permettant d'implémenter les mécanismes d'authentification, limites d'usage et permissions.
- Automatisation et intégration : les API facilitent l'automatisation dans la collecte des données.

En revanche les API comme les autres moyens ont aussi leur limite:

- Seuil limite et usages soumis à certaines restrictions.
- Qualité des données et réutilisabilité : la qualité des données dépend de celui qui les fournit.

Ci dessous nous pouvons lister comme exemple d'API:

- Yahoo Finance API: qui fournit un accès aux données des marchés financiers
- Twitter API: permet l'accès à la vaste collection des tweets et données des utilisateurs
- GoogleMap API: fournit un accès à la localisation en temps réels

Pour illustrer cette section, on va pratiquer un cas de collecte de données sur l'API de Yahoo pour le faire on a besoin des bibliothèques Python suivantes:

- [yfinance](#) (`pip install yfinance`)
- [yahoofinancials](#) (`pip install yahoofinancials`)

La suite se fera en pratique alors rendez-vous dans le dossier workshop pour avoir accès au fichier [collectDataFromAPI.ipynb](#)

## 0.6 Conclusion

Tout au long de ce tutoriel , il etait question de faire un bref rappel partiel de la data engeneering pour que beaucoup aient un apercu de ce qu'est la data engeneering.