



Mohamed Ezer  
GHARIANI, Shayma Jlassi,  
Amal Naamoun,  
Oumayma Mejri, Youssef  
Mechfer

# PROJET DATA SCIENCE

RAPPORT DE PROJET 2

### **1. Compréhension du Problème :**

Notre objectif est de collecter des données relatives aux statistiques d'accidents à partir du site Web <https://www.ins.tn/statistiques/117>. Les informations spécifiques recherchées sont présentes dans plusieurs tables, notamment celles liées à l'évolution du nombre d'accidents par cause, par milieu, et les données sur le nombre d'accidents, de blessés et de tués.

### **2. Collecte de Données :**

Nous utilisons le langage de programmation Python en conjonction avec les bibliothèques requests et BeautifulSoup. Le processus consiste à envoyer une requête HTTP à l'URL spécifiée, puis à extraire les données HTML de la page pour les analyser.

### **3. Préparation des Données :**

Les données ne nécessitent pas de préparation complexe dans ce contexte, car nous avons directement ciblé les tables d'intérêt. La structure des données est préservée sous forme de tableaux et les cellules sont traitées pour extraire le texte.

Code:

```

10 if response.status_code == 200:
11     # Parse the HTML content
12     soup = BeautifulSoup(response.text, 'html.parser')
13
14     # Find elements containing specific data-target attribute
15     data_target_values_of_interest = ["#2", "#3", "#4"] # Update with actual values
16
17     for data_target_value in data_target_values_of_interest:
18         # Locate the element with the specific data-target attribute
19         element = soup.find('a', {'data-target': data_target_value, 'class': 'btn btn-link'})
20
21         if element:
22             # Navigate to the parent table of the current element
23             table = element.find_next('table')
24
25             if table:
26                 # Extract data from the table
27                 table_data = []
28                 for row in table.find_all('tr'):
29                     row_data = [cell.get_text(strip=True) for cell in row.find_all(['th', 'td'])]
30                     table_data.append(row_data)
31
32                 # Write table data to CSV file
33                 title = element.get_text(strip=True)
34                 csv_filename = f'{title}.csv'
35                 with open(csv_filename, 'w', newline='', encoding='utf-8') as csvfile:
36                     csv_writer = csv.writer(csvfile)
37                     csv_writer.writerows(table_data)
38
39                 print(f'Data for "{title}" table has been saved to {csv_filename}')
40             else:
41                 print(f'Table not found for element with data-target "{data_target_value}"')
42         else:
43             print(f'Element with data-target "{data_target_value}" not found on the page.')
44     else:
45         print(f"Error: Unable to retrieve the page (Status Code: {response.status_code})")
46

```