

Projet de Web Scrapping

A parti du site armedforces.eu

Quentin Le Boulch

Table des matières

1.	Introduction	2
2.	Méthodologie : Collecte des données via Web Scrapping.....	2
2.1.	Scraping des pages d'équipements militaires sur armedforces.eu	2
2.1.1.	Organisation et préparation	2
2.1.2.	Scrapping des pages principales et des sous-pages.....	3
2.1.3.	Scraping des pages de chaque pays.....	6
3.	Extraction des données et formation des dataframe	7
3.1.	Processus d'extraction et construction de df_origin_equipment :	7
3.2	Extraction des données et formation du dataframe army_stats_countries	8
3.3	Extraction des données et formation du dataframe df_equipment_details	9
4	Visualisation des données	9
4.1	Visualisation des données militaires mondiales	9
4.2	Analyse des constructeurs d'équipements de défense et des avions multirôles	12
5	Difficultés rencontrées	16
5.1.	Scraping de nombreuses pages et gestion des données.....	16
Annexe :	17

1. Introduction

Dans le cadre de ce projet, j'ai choisi de travailler sur le site « armedforces.eu », qui regroupe une grande quantité d'informations concernant les armées des pays du monde. Ce site fournit des données variées telles que les équipements militaires (terrestres, aériens, navals), les budgets de défense, les effectifs, et des informations sur la puissance nucléaire (entre autres). L'objectif de ce travail est d'extraire ces informations à l'aide de techniques de web scraping en Python, puis de les analyser sous formes graphique pour en tirer des enseignements pertinents.

En nous appuyant sur ces données, nous souhaitons explorer certaines dynamiques globales : quels sont les pays les mieux équipés militairement ? Quelles sont les différences entre les pays en termes de budget défense et de capacités de défense ?

Le site « armedforces.eu » regroupe un large éventail d'informations, mais ces données ne sont pas directement exploitables en l'état. Comment pouvons-nous collecter, organiser et analyser ces données afin de mettre en lumière les pays les plus puissants militairement ? En utilisant des outils comme Python nous cherchons à automatiser la collecte des données, traiter leur organisation et proposer des visualisations claires pour comprendre les tendances et les écarts entre les différentes nations.

2. Méthodologie : Collecte des données via Web Scraping

Pour collecter les données du site armedforces.eu, nous allons faire du web scraping. Le processus repose sur une fonction Python qui va utiliser des expressions régulières des codes sources et une boucle pour automatiser la récupération des données.

2.1. Scraping des pages d'équipements militaires sur armedforces.eu

Pour collecter les données liées aux équipements militaires, nous avons ciblé les pages du site armedforces.eu qui regroupent les informations sur les forces terrestres, aériennes et maritimes. Chaque URL correspond à une catégorie spécifique d'équipement, comme les chars, les avions de combat, ou les navires. (Voir Figure 1)

2.1.1. Organisation et préparation

Une liste d'URLs représentant les différentes catégories d'équipements militaires a été définie. Ces URLs constituent le point de départ pour notre scraping.

```
# Liste des URLs à scraper
urls = [
    'https://armedforces.eu/air_forces/attack',
    'https://armedforces.eu/air_forces/helicopter',
```

```
'https://armedforces.eu/navy/submarine',
'https://armedforces.eu/navy/destroyer',
'https://armedforces.eu/navy/corvette',
'https://armedforces.eu/navy/frigate',
'https://armedforces.eu/navy/aircraftcarrier',
'https://armedforces.eu/land_forces/rocket_artillery',
'https://armedforces.eu/land_forces/self_propelled_artillery',
'https://armedforces.eu/land_forces/afv',
'https://armedforces.eu/land_forces/tanks',
'https://armedforces.eu/air_forces/drone',
'https://armedforces.eu/air_forces/multirole',
'https://armedforces.eu/air_forces/fighter',
]

headers = {
    "User-Agent": "Mozilla/5.0 (X11; Linux x86_64; rv:128.0)
Gecko/20100101 Firefox/128.0"
}
```

Code 1

Les données collectées sont sauvegardées dans un dossier global nommé **Equipement**. Chaque catégorie d'équipement possède un sous-dossier dédié pour stocker les fichiers HTML collectés.

```
if not os.path.exists('Equipement'):
    os.makedirs('Equipement')
```

Code 2

2.1.2. Scrapping des pages principales et des sous-pages

Une fonction **scrape_page** est définie pour gérer les requêtes vers chaque URL de la liste. Cette fonction télécharge la page HTML principale et la sauvegarde dans le sous-dossier correspondant.

```
# Fonction pour scraper une URL et organiser les fichiers dans des sous-
dossiers
def scrape_page(url, folder_name):
    req = requests.get(url, timeout=10, headers=headers)
    if req.status_code != 200:
        print(f'Erreur pour {url}: {req.status_code}')
        return

    # Contenu de la page principale
    contenu = req.text

    equipement_folder = os.path.join('Equipement', folder_name)
    if not os.path.exists(equipement_folder):
        os.makedirs(equipement_folder)

        with open(os.path.join(equipement_folder, f'{folder_name}.html'),
'w', encoding='utf8') as output:
            output.write(contenu)
```

Explication :

- `req = requests.get(url, timeout=10, headers=headers)` : Envoie une requête HTTP GET à l'URL spécifiée avec un timeout de 10 secondes.
- `req.status_code` : Vérifie si la requête a réussi (code 200). Si ce n'est pas le cas, une erreur est affichée.
- `os.makedirs(equipement_folder)` : Crée un sous-dossier pour la catégorie d'équipement si celui-ci n'existe pas déjà.
- `output.write(contenu)` : Sauvegarde le contenu HTML dans un fichier portant le nom de la catégorie.

Ensuite, en analysant le code du contenu HTML, nous utilisons une expression régulière pour identifier les liens (voir **Figure 2**) vers des sous-pages contenant des informations détaillées sur chaque équipement.

```
pattern = r'>\r\n\s*<option value="/(.*)">'
results = re.findall(pattern, contenu)
```

Explication :

- `re.findall(pattern, contenu)` : Cette fonction recherche dans le contenu HTML (contenu) tous les segments de texte qui correspondent au motif spécifié par l'expression régulière. Elle retourne une liste de tous les matchs trouvés.
- L'expression `r'>\r\n\s*<option value="/(.*)">'` :
 - `>` indique la fin d'une balise HTML précédente.

- `\r\n\s*<` capture les espaces de mise en page, y compris les retours à la ligne, qui peuvent varier selon le formatage du code source HTML.
- `<option value="/(.*?)">` cible spécifiquement les balises `<option>` où value contient un chemin relatif vers une sous-page, capturant ce chemin grâce au groupe `(.*?)` qui permet une capture compact des caractères jusqu'à la fin de l'attribut value.

Une fois les liens des sous-pages extraits, ils sont utilisés pour télécharger chaque page individuelle. Ces pages sont également sauvegardées dans le même sous-dossier que la page principale.

```
# Téléchargement et enregistrement des sous-pages
for k in results:
    page_url = f'https://armedforces.eu/{k}'
    print(f'Téléchargement de: {page_url}')
    page_req = requests.get(page_url, timeout=10, headers=headers)
    if page_req.status_code == 200:
        filename = os.path.join(equipement_folder, f'{k.split("/")[-1]}.html')
        with open(filename, 'w', encoding='utf8') as file:
            file.write(page_req.text)
            print(f'Enregistré: {filename}')
    else:
        print(f'Échec pour {page_url} avec statut
```

Explication :

- `page_url = f'https://armedforces.eu/{k}'` : Construit l'URL complète de la sous-page en ajoutant le préfixe du domaine au chemin relatif extrait. Lorsque k est inséré dans `page_url = f'https://armedforces.eu/{k}'`, cela forme l'URL complète vers une sous-page spécifique, comme : https://armedforces.eu/land_forces/tanks/Type_96, qu'on a téléchargé avec le pattern juste au-dessus.

k correspond à la partie de l'URL relative à chaque équipement spécifique listé sur la page principale du site armedforces.eu.

- `page_req.status_code == 200` : Vérifie si la requête a réussi (statut 200 signifie OK).
- `filename = os.path.join(equipement_folder, f'{k.split("/")[-1]}.html')` : Crée un nom de fichier en utilisant le dernier segment de l'URL comme nom de base, garantissant ainsi un nom unique pour chaque sous-page dans le dossier correspondant.
- `with open(filename, 'w', encoding='utf8') as file` : Ouvre un fichier pour écrire le contenu téléchargé, en utilisant l'encodage UTF-8 pour gérer correctement les caractères spéciaux.

Pour automatiser le processus, on boucle sur la liste d'URLs en appelant la fonction `scrape_page` pour chaque URL avec le nom de dossier approprié. Cela permet ainsi le scraping et le stockage de toutes les données.

```
for url in urls:
    # Nom du sous-dossier basé sur le dernier segment de l'URL
    folder_name = url.split("/")[-1]
    scrape_page(url, folder_name)
```

Explication :

- `url.split("/")[-1]` : Extrait le dernier segment de l'URL pour utiliser comme nom de dossier (qui correspond à la catégorie de l'équipement).
- `scrape_page(url, folder_name)` : Appelle la fonction `scrape_page` qui gèrera le téléchargement de la page principale et de toutes ses sous-pages associées. Cela permettant d'organiser les fichiers dans le dossier correspondant.

Ainsi, ce processus assure une collecte efficace et structurée des données sur les équipements militaires du site `armedforces.eu`.

2.1.3. Scraping des pages de chaque pays

Nous procédons au scraping des pages de chaque pays de la même manière que pour les équipements militaires, avec quelques ajustements spécifiques :

Nous commençons par télécharger le contenu de la page principale qui liste tous les pays sur le site `armedforces.eu`. Après avoir obtenu cette page nous utilisons une expression régulière spécifiquement conçue pour cette tâche afin d'extraire les URL des pages individuelles de chaque pays. Cette expression régulière est ajustée pour correspondre aux balises HTML qui contiennent les liens des pays.

Téléchargement et sauvegarde des pages de pays :

Nous créons un dossier nommé 'countries' pour centraliser et organiser les fichiers HTML de chaque pays. Une fois ce dossier établi, nous procédons au téléchargement de chaque page de pays listée dans les liens extraits, puis sauvegardons chaque contenu de page dans ce dossier.

```
if not os.path.exists('countries'):
    os.makedirs('countries')

# Téléchargement des pages des pays
for country_url in country_links:
    country_name = country_url.split('/')[-1]
```

```

print(f'Téléchargement de la page pour {country_name} depuis
{country_url}')

country_response = requests.get(country_url, headers={"User-Agent":
"Mozilla/5.0"})
if country_response.status_code == 200:
    file_path = os.path.join('countries', f'{country_name}.html')
    with open(file_path, 'w', encoding='utf-8') as file:
        file.write(country_response.text)
    print(f'Sauvegardé: {file_path}')
else:
    print(f'Échec pour {country_name}. Statut:
{country_response.status_code}')

```

Code 3

Explication groupée :

Ce code gère le téléchargement et la sauvegarde des pages web des pays depuis le site armedforces.eu :

- **Vérification et création du dossier 'countries' :** Il vérifie d'abord si le dossier 'countries' existe, et si ce n'est pas le cas, il le crée pour y stocker les fichiers HTML des pays.
- **Boucle sur les URLs des pays :** Pour chaque URL dans country_links, il extrait le nom du pays du dernier segment de l'URL et annonce le début du téléchargement.
- **Gestion de la réponse :**
 - ❖ Si la requête est réussie (statut 200), le contenu HTML de la page est sauvegardé dans un fichier nommé d'après le pays dans le dossier 'countries' . Un message confirme la sauvegarde.
 - ❖ Si la requête échoue, un message d'erreur est affiché contenant le nom du pays et le code d'erreur.

Ce processus permet de collecter systématiquement les informations sur les forces militaires par pays, en les organisant de manière structurée pour une utilisation ou une analyse ultérieure.

3. Extraction des données et formation des dataframe

3.1. Processus d'extraction et construction de df_origin_equipment :

Nous commençons par définir une fonction d'extraction pour isoler l'origine de chaque équipement à partir des fichiers HTML. Cette fonction utilise une expression régulière pour identifier précisément où se trouve cette information. Puis, nous parcourons le dossier parent "Equipment", examinant chaque sous-dossier et chaque fichier HTML pour extraire cette origine. Pour chaque fichier, nous créons un

dictionnaire contenant la catégorie de l'équipement (nom du sous-dossier), le nom de l'équipement (basé sur le nom du fichier sans extension) et son origine.

Ensuite, nous compilons toutes ces informations dans une liste de dictionnaires. Cette liste est convertie en un DataFrame Pandas nommé `df_origin_equipment`. Nous procédons à un nettoyage des noms des équipements en remplaçant les underscores par des espaces afin d'améliorer la lisibilité des données.

Pour conclure, nous affichons ce DataFrame pour une vérification immédiate et le sauvegardons au format CSV. Lors de la sauvegarde, nous veillons à exclure l'index du DataFrame et à utiliser un encodage UTF-8 pour assurer la compatibilité avec les caractères spéciaux.

Ce processus nous permet de structurer des données initialement semi-structurées dans un format prêt pour l'analyse ou l'archivage.

Le dataframe `df_origin_equipment` (voir Figure 3) que nous obtenons présente des informations sur les types d'équipements militaires ainsi leur pays d'origine et catégorie.

3.2 Extraction des données et formation du dataframe `army_stats_countries`

Ce processus implique l'initialisation de liste/tuple pour différents types de données militaires, incluant la population (`Population`), le budget militaire (`Military Budget (USD billion)`), les armes nucléaires (`Nuclear Weapons`), le personnel militaire actif (`Active Personnel (thousands)`), la réserve militaire (`Military Reserve (thousands)`), ainsi que divers types d'équipements militaires comme les chars (`Tanks`), les véhicules blindés de combat (`AFV`), l'artillerie (`Artillery, Self-propelled guns, Rocket artillery`), les aéronefs (`Total aircraft, Fighters, Multirole, Attack, Helicopters, UCAV (combat drone)`), et les navires (`Total naval, Aircraftcarriers, Destroyers, Frigates, Corvettes, Submarines`).

Pour l'extraction de ces données, nous utilisons une méthode basée sur des fichiers HTML des pays. Une fonction générique `extract_data` est définie pour utiliser des expressions régulières (regex) spécifiques afin d'extraire les informations de chaque catégorie. Ensuite, chaque fichier HTML est parcouru, les données correspondantes sont extraites et stockées dans un dictionnaire pour chaque pays. Ces dictionnaires sont ensuite compilés dans un DataFrame Pandas nommé `army_stats_countries` pour faciliter l'analyse. Les noms des pays sont nettoyés en remplaçant les underscores par des espaces pour améliorer la lisibilité avant l'affichage final du DataFrame (voir Figure 4).

3.3 Extraction des données et formation du dataframe df_equipement_details

Nous initialisons un dictionnaire `equipment_data` pour stocker les informations sur les équipements militaires extraites des fichiers HTML dans le dossier "Équipement". Une fonction `extract_equipement_data` utilise des expressions régulières pour extraire les détails des équipements (pays, quantité, version). Chaque fichier HTML est traité, les données sont ajoutées au dictionnaire, puis converties en un DataFrame Pandas (`df_equipement_details`) contenant les colonnes `Equipment`, `Country`, `Quantity`, et `Version`. Le DataFrame est affiché et peut être sauvegardé en CSV (voir Figure 5).

4 Visualisation des données

Après avoir collecté et structuré nos données sur les forces militaires mondiales, nous nous sommes penchés sur leur visualisation pour mieux comprendre la distribution des équipements militaires, les budgets.

Packages utilisés :

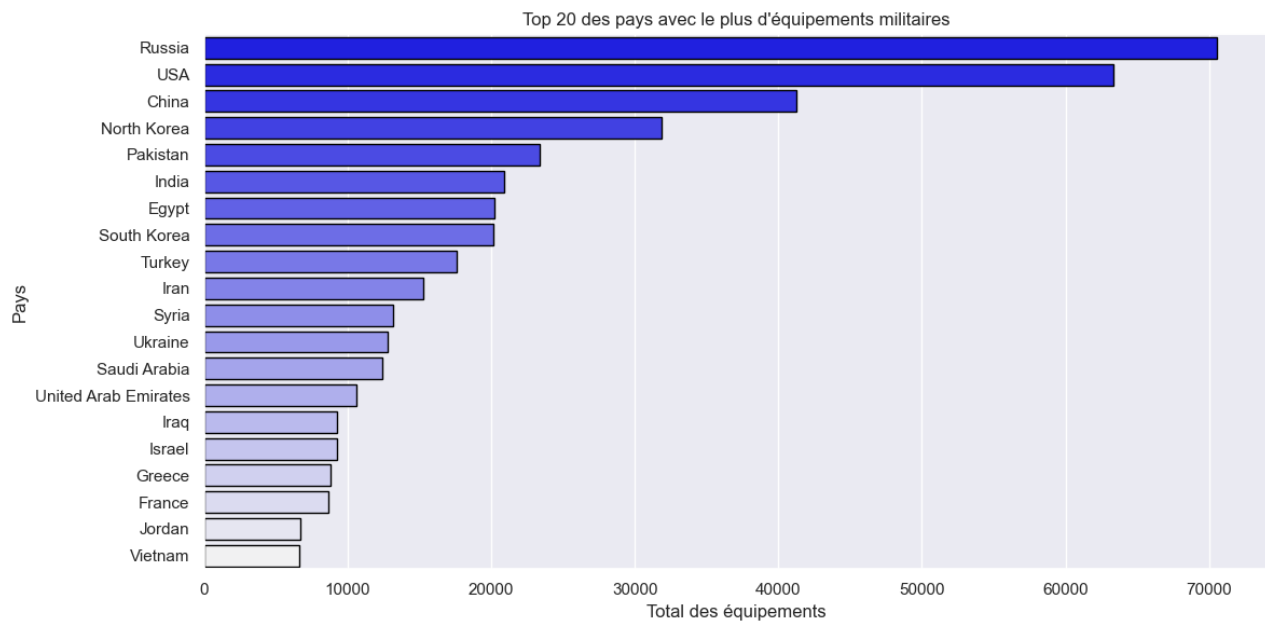
- **Pandas (pd)** : Utilisé pour nettoyer et convertir les données en format numérique, puis pour calculer la somme totale des équipements par pays (`Total equipments`).
- **Matplotlib (plt)** : Fournit la base pour la création de graphiques, ici utilisé pour définir la taille de la figure et afficher le graphique.
- **Seaborn (sns)** : Construit sur Matplotlib pour offrir des visualisations statistiques attrayantes. Ici, il est utilisé pour créer un graphique en barres avec une palette de couleurs personnalisée (`light_palette`), pour styliser le graphique (`set`), et pour ajouter des détails comme les contours noirs autour des barres.
- **NumPy (np)** : Bien que mentionné dans l'importation, il n'est pas directement utilisé dans ce snippet.

4.1 Visualisation des données militaires mondiales

En explorant la distribution des forces militaires à travers le monde, nous avons créé trois histogrammes pour offrir une vue d'ensemble détaillée et comparative de la puissance militaire des nations.

1^{er} graphique :

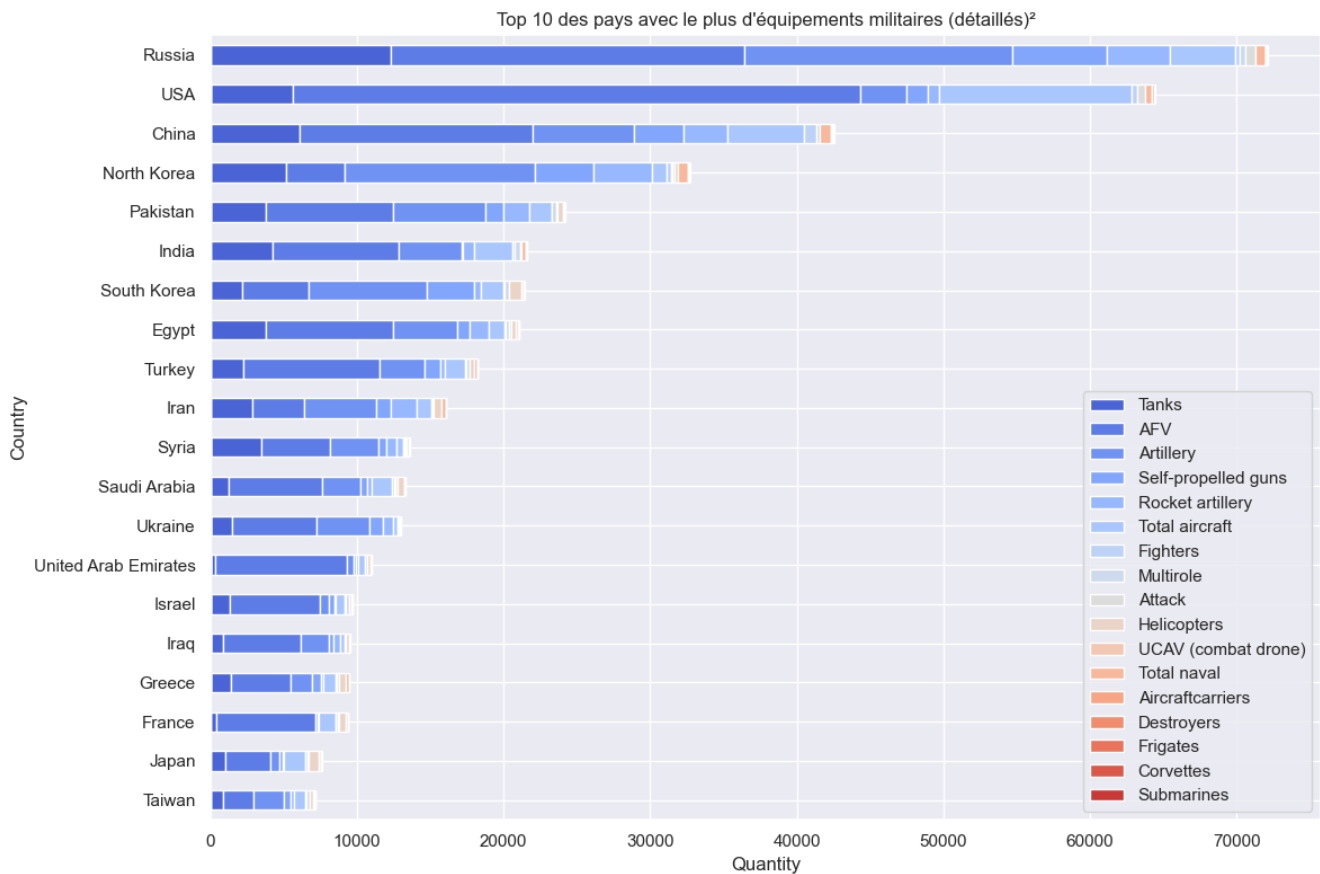
Graphique 1 : **Top 20 des pays avec le plus d'équipements militaires**



Cet histogramme nous montre que la Russie et les États-Unis mènent la course avec un total d'équipements militaires dépassant d'assez loin les autres nations. La Chine, la Corée du Nord, et le Pakistan suivent, mais avec un écart significatif. Ce graphique souligne la course aux armements à l'échelle mondiale reflétant à la fois leurs ambitions géopolitiques et/ou leurs préoccupations sécuritaire car nous avons beaucoup de pays en guerre ou en conflit majeur (Pakistan, Inde, Israël, Iran, Ukraine etc..)

2^{ème} graphique :

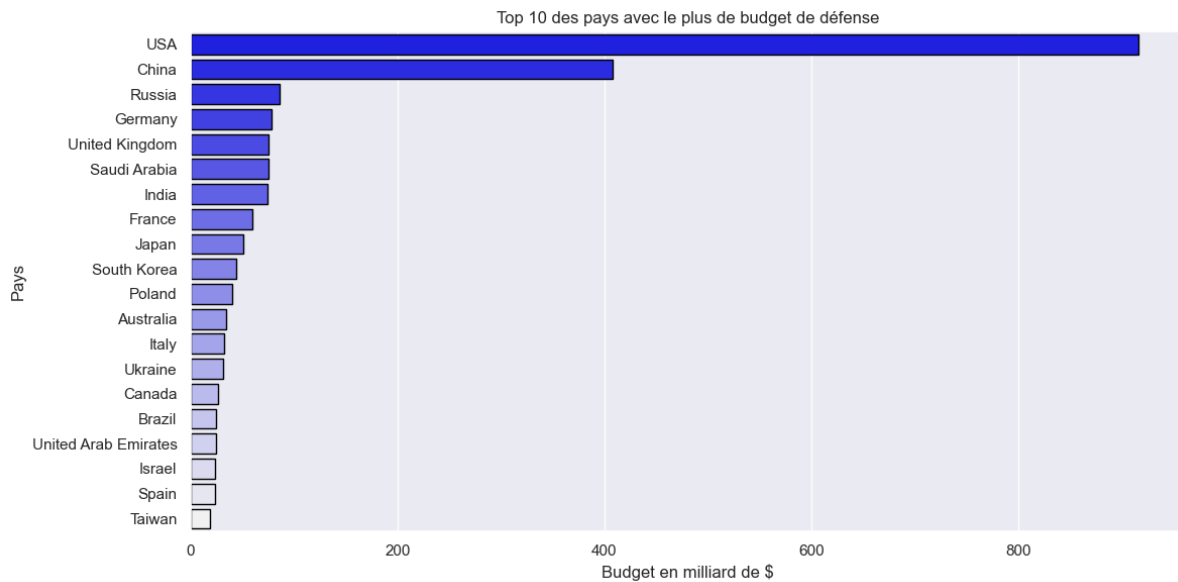
Graphique 2 : **Top 20 des pays avec le plus d'équipements militaires (détaillé):**



En allant plus loin, ce deuxième histogramme décompose les équipements militaires par catégorie pour chaque pays. On observe que la Russie et les USA ne dominent pas seulement en nombre total mais aussi dans une variété d'équipements, des chars aux sous-marins. Cette diversité montre leur capacité à opérer sur plusieurs fronts, tant terrestres que maritimes et aériens. On voit que la grande majorité des équipements des différentes nations sont terrestres avec des tanks, AFV, de l'artillerie, Self-propelled guns. On ne distingue pas ou à peine sur ce graphique les équipements navales ou aériens car ils sont en proportion beaucoup moins nombreux.

3^{ème} graphique :

Graphique 3 : Top 20 des pays avec le plus de budget de défense



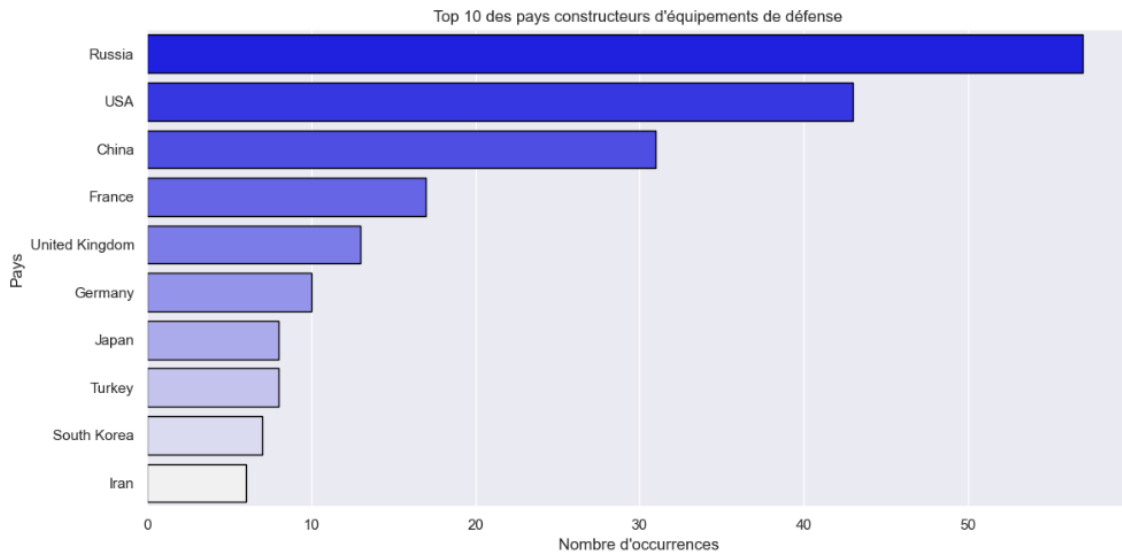
Enfin, en examinant le budget de défense, l'histogramme révèle que les États-Unis dépensent bien plus que n'importe quel autre pays, avec un budget dépassant les 800 milliards de dollars. La Chine et la Russie suivent, mais à une échelle nettement inférieure. Ce graphique met en lumière l'investissement financier colossal des États-Unis dans la défense, ce qui pourrait expliquer leur position dominante dans les deux premiers graphiques. Ce graphique souligne aussi des différences dans le classement avec les graphiques précédents. Le nombre d'équipements par pays, montrant bien dans le premier classement qu'elles sont les pays qui sont vraiment en conflit majeur.

4.2 Analyse des constructeurs d'équipements de défense et des avions multirôles

Poursuivant notre exploration des capacités militaires mondiales, nous avons examiné deux aspects cruciaux : la production d'équipements de défense et la quantité d'avions multirôles en service.

4^{ème} graphique :

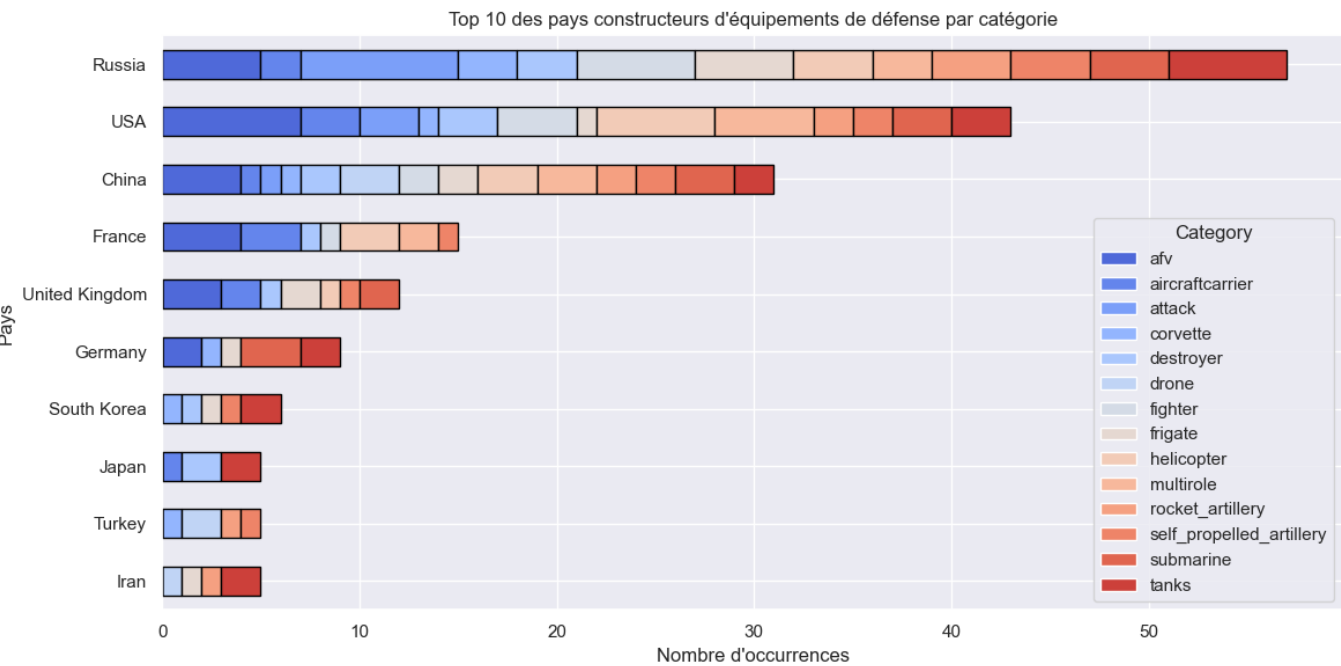
Graphique 4 : Top 10 des pays constructeurs d'équipements de défense



Ce graphique nous révèle que la Russie et les États-Unis sont non seulement des géants en termes de quantité d'équipements militaires mais aussi des leaders dans leur construction. La Russie, en particulier, se distingue avec une production massive, illustrant sa position de puissance militaire autonome avec une capacité de production importante. Les États-Unis, bien que légèrement en retrait, montrent une diversité dans la production militaire. Ce graphique met en lumière l'importance de l'industrie de défense dans la stratégie militaire globale de ces nations.

5^{ème} graphique :

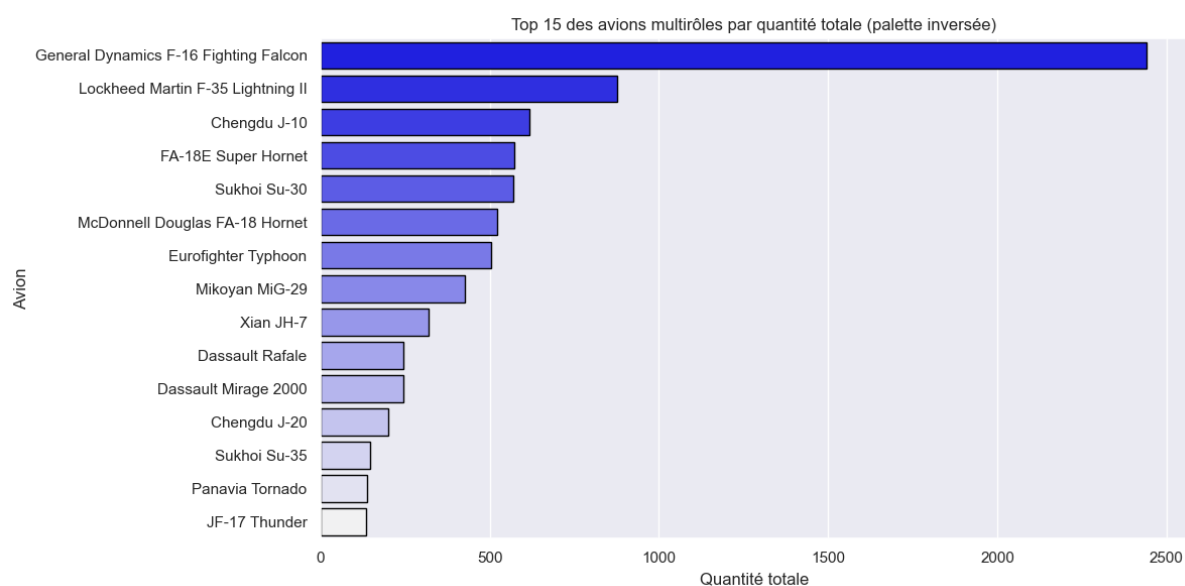
Graphique 5 : Top 10 des pays constructeurs d'équipements de défense par catégorie



En allant plus en détail, ce deuxième histogramme montre comment ces pays répartissent leur production militaire à travers différentes catégories d'équipements (par exemple le nombre de modèle de tank ou de afv différents, on distingue donc les modèles). La Russie et les USA dominent à nouveau, mais ici, nous voyons une spécialisation : la Russie excelle dans la production de chars (tanks) et de sous-marins par exemple. Les États-Unis montrent une large gamme de production, y compris des porte-avions (aircraftcarrier) et des drones. Cette diversité dans la production met en évidence comment chaque pays peut influencer les capacités militaires mondiales à travers des technologies spécifiques.

6^{ème} graphique :

Graphique 6 : Top 15 des avions multirôles par quantité totale



Pour approfondir nous sommes focalisés sur les avions multirôles, ce graphique nous montre que le General Dynamics F-16 Fighting Falcon (avion Américain) est de loin l'avion multirôle le plus produit, suivi par le Lockheed Martin F-35 Lightning II. Cette prédominance du F-16 souligne son rôle crucial dans de nombreuses forces aériennes à travers le monde, offrant une flexibilité opérationnelle inégalée. La présence significative du F-35, malgré sa production plus récente, indique un investissement massif dans la technologie de pointe et la modernisation des flottes aériennes.

Enfin, nous avons sélectionné un multirôle par pays pour observer quels pays possèdent chacun de ces avions multirôles. Cela nous permet de visualiser et d'estimer les alliances de chaque pays à travers les fournisseurs et acheteurs des avions (voir **Figure 6, Figure 7, Figure 8, Figure 9, Figure 10, Figure 11**)

Par exemple, la France vend son Rafale à l'Égypte, le Qatar, l'Inde et la Grèce (dernières données sur le site armedforces.eu).

Les États-Unis exportent le F-35 au Royaume-Uni, en Australie, au Danemark, en Israël, en Italie, au Japon, aux Pays-Bas, en Norvège et en Corée du Sud.

La Russie, quant à elle, vend le Mikoyan MiG-29 à la Syrie, l'Ukraine (bien que la maintenance ne soit probablement pas assurée dans le contexte actuel), l'Azerbaïdjan, la Biélorussie, l'Égypte, l'Inde, l'Iran, la Corée du Nord et la Pologne.

5 Difficultés rencontrées

5.1. Scraping de nombreuses pages et gestion des données

Le site armedforces.eu offre une vaste collection de pages détaillant les équipements militaires par catégorie (chars, avions, navires, etc.) ainsi que des pages spécifiques pour chaque pays. Scraper manuellement ces données aurait été une tâche fastidieuse et peu efficace. De plus, gérer les fichiers HTML téléchargés et les organiser dans des dossiers spécifiques nécessitait une approche rigoureuse.

Pour automatiser ce processus, une fonction Python, `scrape_page`, a été développée (voir Code 3). Cette fonction prend en entrée une URL et un nom de dossier. Elle télécharge la page principale, puis identifie et extrait toutes les sous-pages associées. Ces fichiers sont ensuite sauvegardés dans des sous-dossiers spécifiques, organisés par catégorie d'équipement.

Ce projet a été grandement facilité par la bonne structuration du site. Un site moins organisé aurait rendu cette tâche bien plus complexe, nécessitant davantage d'efforts pour concevoir la logique d'extraction.

5.2 Gestion du format des données et nettoyage

Lors de l'extraction des données à partir des fichiers HTML, nous avons rapidement constaté que les données étaient souvent stockées sous forme de chaînes de caractères (strings), ce qui posait problème lors des calculs ou des analyses. Par exemple, les budgets militaires étaient souvent accompagnés d'unités comme " `billion`" ou "\$", et les quantités d'équipements pouvaient inclure des virgules ou d'autres caractères non numériques. Cela rendait impossible l'exécution de calculs directs, comme des sommes ou des moyennes, sans un nettoyage préalable des données.

Pour résoudre ce problème, nous avons dû convertir les données en formats numériques appropriés (entiers ou flottants) en supprimant les caractères superflus. Par exemple, pour les budgets militaires, nous avons utilisé des expressions régulières pour supprimer les unités et convertir les valeurs en nombres flottants. De même, pour les quantités d'équipements, nous avons supprimé les virgules et converti les chaînes de caractères en entiers.

Annexe :

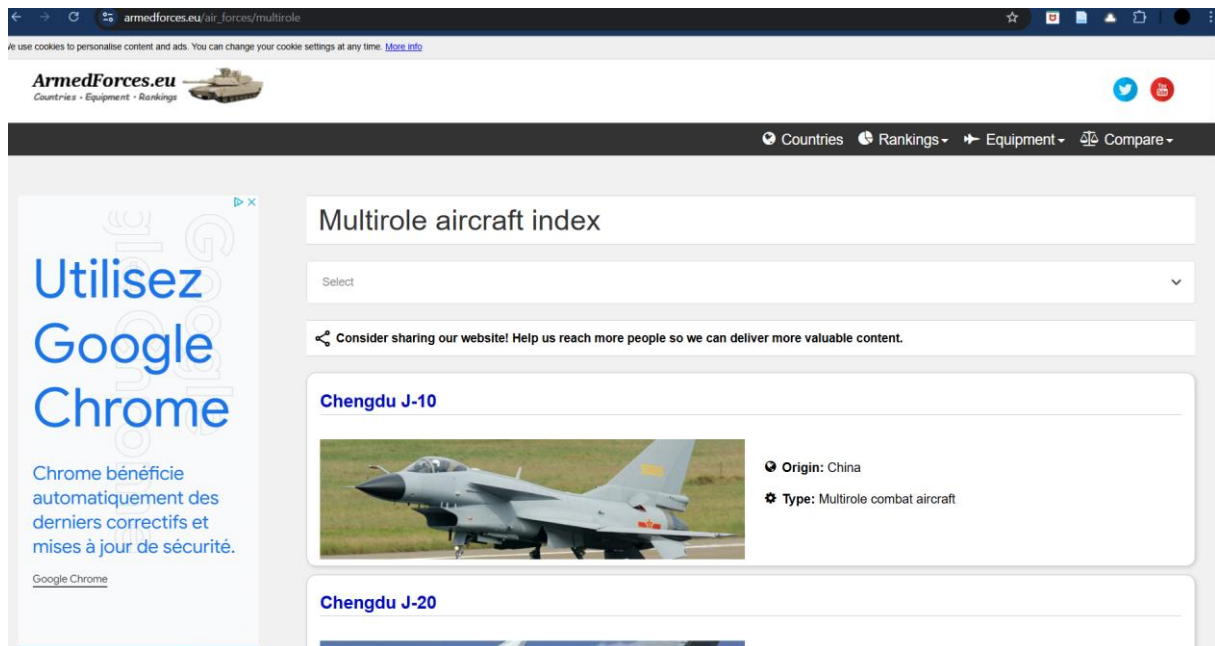


Figure 1 : Exemple Page équipement à télécharger

```
<option value="/land_forces/tanks/Zulfiqar">Zulfiqar</option>
</optgroup>
<optgroup label="Israel">
  <option value="/land_forces/tanks/Merkava_Mark_III">Merkava Mark III</option>
  <option value="/land_forces/tanks/Merkava_Mark_IV">Merkava Mark IV</option>
</optgroup>
<optgroup label="Italy">
  <option value="/land_forces/tanks/Ariete">Ariete</option>
</optgroup>
<optgroup label="Japan">
  <option value="/land_forces/tanks/Type_10">Type 10</option>
  <option value="/land_forces/tanks/Type_90_Ky%C5%AB-maru">Type 90 Kyū-maru</option>
</optgroup>
<optgroup label="North Korea">
  <option value="/land_forces/tanks/Pokpung-ho">Pokpung-ho</option>
</optgroup>
<optgroup label="Pakistan/China">
  <option value="/land_forces/tanks/Al-Khalid">Al-Khalid</option>
  <option value="/land_forces/tanks/Al-Zarrar">Al-Zarrar</option>
</optgroup>
<optgroup label="Poland">
  <option value="/land_forces/tanks/PT-91_Twardy">PT-91 Twardy</option>
</optgroup>
<optgroup label="Romania">
  <option value="/land_forces/tanks/TR-85">TR-85</option>
</optgroup>
```

Figure 2 : Extrait code source de la page des équipements de tanks

Ici l'élément que nous allons chercher, est encadrer en rouge et nous servira pour former les liens des pages de chaque équipement à télécharger

	Category	Equipment	Origin
0	afv	ABC-79M	Romania
1	afv	afv	None
2	afv	AMX 10 RC	France
3	afv	APC Talha	Pakistan
4	afv	ASCOD	Austria/Spain
5	afv	Bionix	Singapore
6	afv	BMP-1	Russia
7	afv	BMP-2	Russia
8	afv	BMP-3	Russia
9	afv	BMR	Spain
10	afv	Boxer	Germany/Netherlands
11	afv	BTR-80	Russia
12	afv	Bushmaster Protected Mobility Vehicle	Australia
13	afv	Cadillac Gage Commando	USA
14	afv	Casspir	South Africa
15	afv	Centauro	Italy
16	afv	CM-32 Yunpao	Taiwan (Republic of China)
17	afv	Combat Vehicle 90 (CV90)	Sweden
18	afv	Cougar	USA
19	afv	EE-9 Cascavel	Brazil

Figure 3 : df_origin_equipment

army_stats_countries

	Country	Population	Military Budget (USD billion)	Nuclear Weapons	Active Personnel (thousands)	Military Reserve (thousands)	Tanks	AFV	Artillery	Self-propelled guns	Rocket artill
0	Armenia	3,000,756	1.5 billion \$	0 warheads	70,600	210,000	269	610	236	38	103
1	Australia	26,037,300	34.4 billion \$	0 warheads	60,330	29,740	73	1,582	229	0	0
2	Austria	9,027,999	3.2 billion \$	0 warheads	23,000	125,600	56	215	30	30	0
3	Azerbaijan	10,353,296	3.8 billion \$	0 warheads	126,400	330,000	940	2,363	909	267	323
4	Belarus	9,255,524	0.9 billion \$	0 warheads	63,000	344,750	500	1,831	771	375	246
5	Brazil	217,240,060	25 billion \$	0 warheads	334,500	1,340,000	469	1,696	760	136	78
6	Canada	39,292,355	27 billion \$	0 warheads	68,000	27,000	74	1,678	159	0	0
7	China	1,411,750,000	408 billion \$	260 warheads	2,035,000	2,000,000	6,100	15,880	6,890	3,390	3,000
8	Denmark	5,982,117	5 billion \$	warheads	21,000	63,000	44	621	27	19	8
9	Egypt	107,770,524	6 billion \$	0 warheads	438,500	479,000	3,789	8,699	4,328	889	1,250
10	Finland	5,566,000	6.7 billion \$	0 warheads	24,000	870,000	239	2,020	821	97	62
11	France	68,042,591	60 billion \$	300 warheads	270,000	141,050	406	6,748	119	96	11
12	Germany	83,695,430	78 billion \$	0 warheads	183,638	34,600	295	3,752	409	134	33
13	Greece	10,432,481	9 billion \$	0 warheads	107,600	221,600	1,365	4,126	1,469	588	152
14	Hungary	9,689,000	5.2 billion \$	0 warheads	37,650	20,000	176	720	298	3	0
15	India	1,375,586,000	74 billion \$	150 warheads	1,400,000	1,155,000	4,215	8,600	4,339	100	700

Figure 4 : dataframe army_stats_countries

	Equipment	Country	Quantity	Version
0	ABC-79M	Romania	406	TABC-79
1	AMX_10_RC	France	210	AMX-10 RC
2	AMX_10_RC	Qatar	12	AMX-10RC
3	APC_Talha	Pakistan	2000	APC Talha
4	ASCOD	Austria	112	ASCOD
5	ASCOD	Spain	261	Pizarro
6	Bionix	Singapore	800	Bionix II
7	BMP-1	Armenia	80	BMP-1
8	BMP-1	Azerbaijan	120	BMP-1
9	BMP-1	Belarus	132	BMP-1
10	BMP-1	China	600	Type 86
11	BMP-1	China	650	Type 86A
12	BMP-1	Egypt	200	BMP-1
13	BMP-1	Greece	184	BMP-1P Ost
14	BMP-1	Iran	450	Boragh/BMP-1
15	BMP-1	Iraq	600	BMP-1

Figure 5 : df_equipment_details

Répartition des quantités pour Dassault Rafale

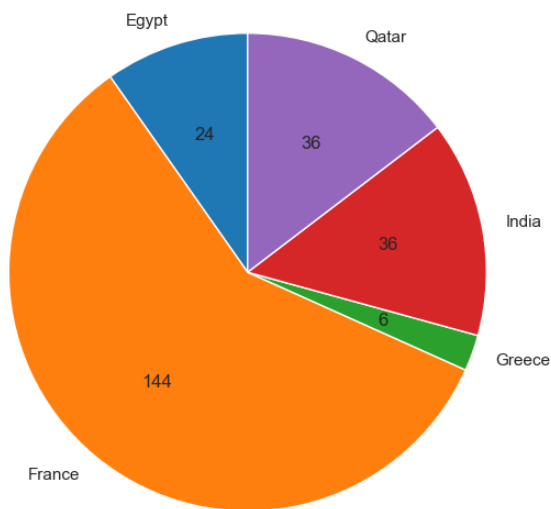


Figure 6: Répartition des quantités pour Rafale Dassault par

Répartition des quantités pour Eurofighter Typhoon

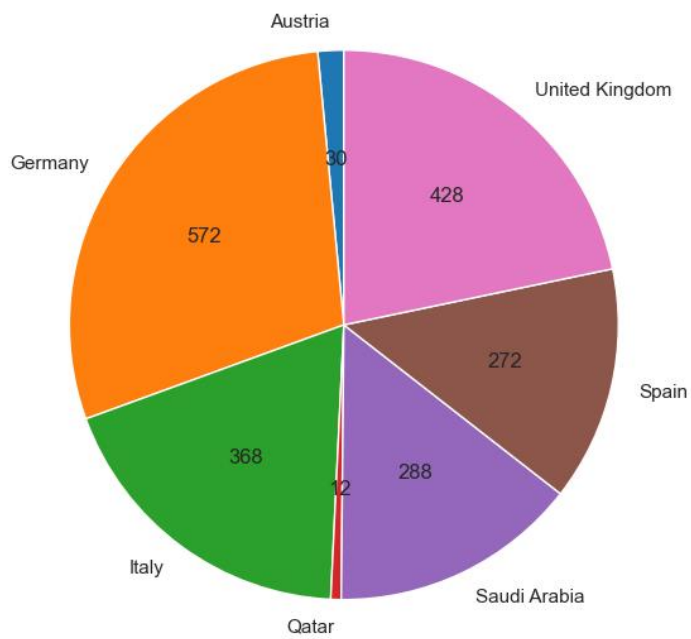


Figure 7 : Répartition des quantités pour Eurofighter Typhoon par Pays

Répartition des quantités pour General Dynamics F-16 Fighting Falcon

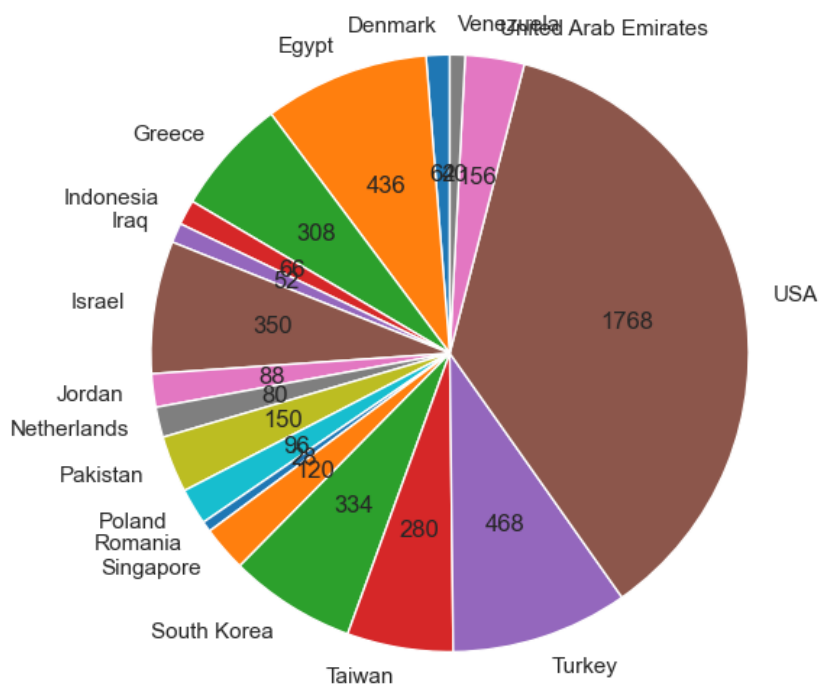


Figure 8 : Répartition des quantités pour le F-16 par pays

Répartition des quantités pour Chengdu J-10

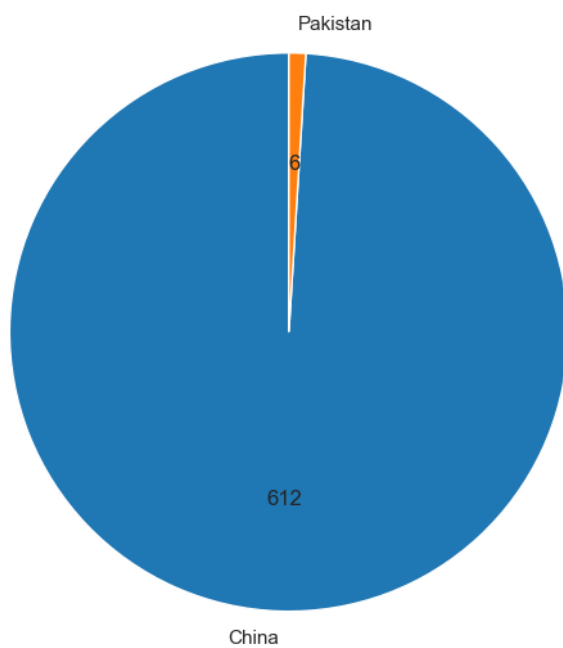


Figure 9 : Répartition des quantités pour Chengdu J-10 par pays

Répartition des quantités pour Lockheed Martin F-35 Lightning II

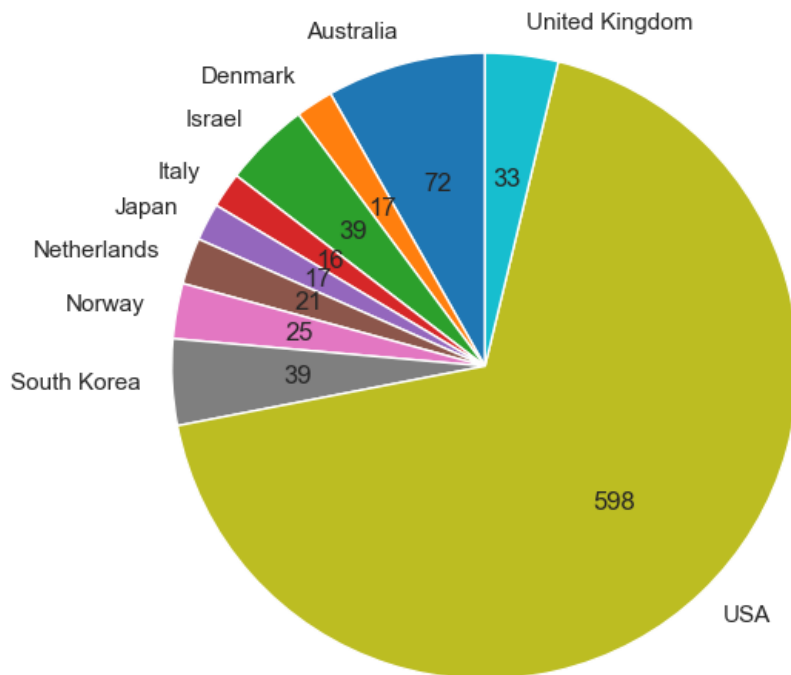


Figure 10 : Répartition des quantités pour le F-35 par pays

Répartition des quantités pour Mikoyan MiG-29

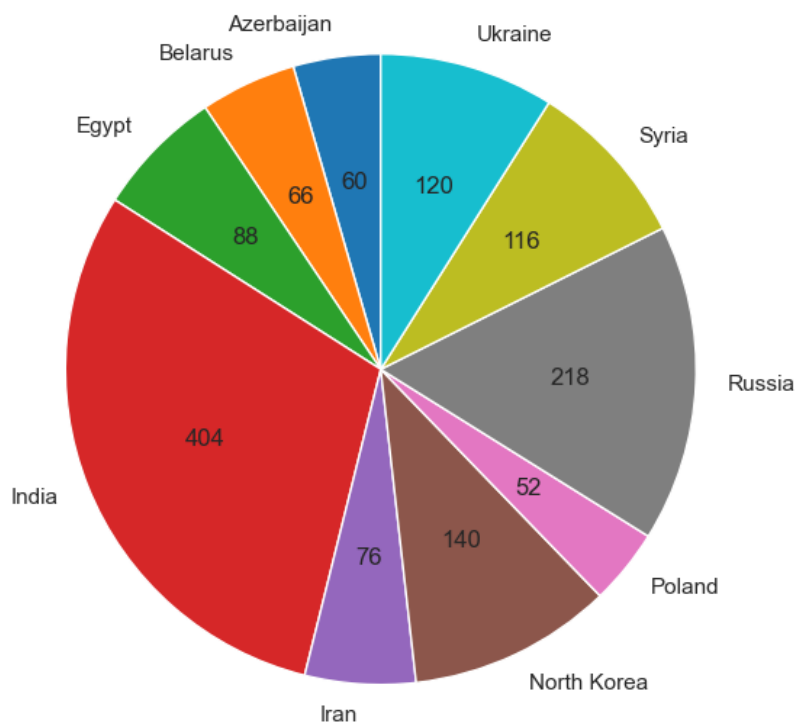


Figure 11 : Répartition des quantités pour Mikoyan MiG-29 par pays