

Le lien entre la hausse de température des océans et l'augmentation de la puissance des ouragans



Introduction.....	2
I. Trouver des datasets.....	3
A. Le dataset des ouragans	3
B. Le dataset des ouragans	5
II. Utilisation des datasets et premières analyses.....	5
III. Analyse du lien entre nos datasets	7
III. Études approfondies	9
1- Trouver l'ouragan le plus fort, sa durée, le mois pendant lequel il est survenu	9
2- Trouver l'ouragan le moins fort, sa durée, le mois pendant lequel il est survenu	10
3- Généraliser cela en montrant que les ouragans ont tendance à se produire à une certaine période de l'année et qu'il est rare qu'ils se produisent à une autre certaine période de l'année	10
3-1) Quelle est la durée moyenne des ouragans ?	10
3-2) Quelle est la période de l'année la plus propice aux ouragans ?.....	11
Sources	12

Introduction

Dans le cadre de l'UE Bio-Informatique, où l'on s'intéresse aux changements climatiques en tout genre, nous nous sommes intéressés au lien de causalité entre la hausse des températures des océans, la puissance des ouragans qui frappent l'océan Atlantique et la fréquence des tempêtes tropicales.

Avant de commencer, faisons un point vocabulaire, en précisant la différence entre un **ouragan**, un **cyclone** et un **typhon**. D'après le site de National Geographic [3], les 3 proviennent du même phénomène météorologique ; la seule différence est dans la provenance de la tempête, c'est-à-dire son origine géographique. Les ouragans se situent dans les océans Pacifique et Atlantique Nord, les typhons sont au nord-ouest du Pacifique et les cyclones sont dans l'océan Indien, et sont divisés en trois sous catégories également selon sa localisation (nord, sud-ouest, ...). Les ouragans, typhons et cyclones sont considérés comme tel à partir de **119 km/h**. Un ouragan est considéré "intense" à partir de 179 km/h.

Faisons également un point sur ce qu'est la température des océans. Celle-ci correspond à la chaleur mesurée à la surface. Elle est influencée par divers facteurs tels que le rayonnement solaire, les courants marins et les échanges thermiques avec l'atmosphère. Elle est mesurée par des satellites qui vérifient la quantité d'énergie rejetée par l'océan à différentes longueurs d'ondes. Ensuite, les programmes informatiques combinent les données des satellites avec les données de navires et de bouées, tout en incorporant l'information des cartes de glace de mer. Depuis plusieurs décennies, les relevés montrent une augmentation progressive de cette température, un phénomène largement attribué au réchauffement climatique.

Nous avons choisi de nous restreindre à l'Atlantique Nord pour plusieurs raisons. Tout d'abord, cette région est particulièrement bien documentée, avec des données historiques plus précises et détaillées que dans d'autres bassins océaniques. Ensuite, l'Atlantique Nord est directement concerné par les ouragans qui menacent régulièrement les côtes des États-Unis, des Caraïbes et du Golfe du Mexique, avec des impacts humains et économiques majeurs. Enfin, cette restriction nous permet d'affiner notre analyse en nous concentrant sur un bassin spécifique, ce qui facilite l'interprétation des tendances et la mise en évidence de corrélations entre la température de l'océan et l'évolution des tempêtes tropicales.

Nous nous sommes intéressés à ce sujet pour plusieurs raisons. Tout d'abord, l'un des membres de notre groupe est personnellement touché par ce sujet, ayant vécu toute sa vie en Guadeloupe. Il nous a informé des conséquences dévastatrices de ce type de phénomène naturel, ce qui a suscité notre intérêt à étudier le rôle du dérèglement climatique dans tout cela. Puis, à la suite de la découverte d'un

article intitulé *"Improved simulation of 19th- and 20th-century North Atlantic hurricane frequency after correcting historical sea surface temperatures"* [1], notre curiosité s'est étendue. Nous n'avons pas repris de données, de graphiques ou de code de cet article, mais nous nous sommes inspirés de l'idée sous-jacente, qui est de montrer un lien concret entre la hausse des températures de l'océan, et la hausse de puissance et de fréquence des ouragans depuis 1980. Ainsi, nous nous sommes focalisés sur les tempêtes de l'océan Atlantique nord.

Mieux expliquer pq on s'est restreint à l'atlantique nord?

I. Trouver des datasets

A. Le dataset des ouragans

La première étape a été de trouver un dataset avec tous les ouragans depuis un nombre d'années assez conséquent afin de pouvoir effectuer une analyse pertinente de l'impact du dérèglement climatique sur la force des ouragans de l'océan Atlantique.

Nous avons trouvé un dataset sur le site de la NOAA (The National Oceanic and Atmospheric Administration - gouvernement américain) [2]. Ce jeu de données est la base de données des ouragans de l'Atlantique de la NOAA. Les données comprennent les positions et les attributs des tempêtes de 1975 à 2022. Les tempêtes à partir de 1979 sont mesurées toutes les six heures pendant la durée de vie de la tempête. Certaines données manquent pour les tempêtes des années précédentes.

Afin de faciliter notre analyse et de minimiser le coût de notre base de données, nous avons procédé à un tri du dataset.

Tout d'abord, nous avons supprimé certaines colonnes qui n'étaient pas utiles pour notre analyse. Notamment le diamètre de force et la pression des tempêtes.

Puis nous avons décidé de plus se focaliser sur les ouragans du dataset, et moins sur les tempêtes tropicales, dépressions tropicales, etc. Il y a deux raisons à cela.

- Premièrement, au vu de la quantité de données et de la contrainte de temps, nous nous sommes dit que notre analyse serait plus détaillée et intéressante en se limitant à un seul type de tempête.
- Deuxièmement, il y avait pas mal de données manquantes pour les autres types de tempêtes. Ainsi, pour les tempêtes tropicales, nous avons seulement mesuré la hausse de fréquence, et non la puissance du vent.

Ensuite, les ouragans sont mesurés toutes les 6 heures et cela fait beaucoup de données à analyser qui ne sont pas forcément pertinentes pour notre étude. Ainsi, nous avons décidé de seulement garder les moments de force du vent minimum et maximum pour chaque ouragan, généralisant notre étude aux extrêmes. Il est important de noter que la vitesse minimum d'un ouragan est fixée à 65 km/h. Cela nous permet de se focaliser sur les vitesses max.

Une *difficulté* que nous avons rencontrée ici est qu'au fil des années, des ouragans peuvent avoir le même nom. Par exemple, pour les ouragans de l'océan l'Atlantique, il y a une liste de noms sur 6 ans. En autres mots, une liste est répétée tous les 6 ans. Dans le cas particulier d'un ouragan très dévastateur, le nom est retiré de la liste. Ainsi, nous avons dû faire attention à regrouper les ouragans du dataset par année, puis par nom, et enfin par puissance minimum ou maximum du vent, afin de conserver toutes les données qui nous intéressent.

À la suite du tri du dataset, nous nous sommes retrouvés avec deux datasets min et max où l'on retrouve 319 ouragans, tous situés dans l'océan Atlantique Nord.

Notre analyse se situe donc autour du dataset max (car le min est obsolète).

1er Dataset: MAX

	name	year	month	day	hour	lat	long	status	category	wind
0	Blanche	1975	7	27	18	37.9	-68.0	hurricane	1.0	75
1	Caroline	1975	8	31	0	24.0	-97.0	hurricane	3.0	100
2	Doris	1975	9	2	6	36.4	-44.5	hurricane	2.0	95
3	Eloise	1975	9	23	12	30.2	-86.3	hurricane	3.0	110
4	Faye	1975	9	27	0	32.7	-64.2	hurricane	2.0	90
..
314	Fiona	2022	9	21	6	23.5	-71.8	hurricane	4.0	120
315	Ian	2022	9	28	12	26.0	-82.7	hurricane	5.0	140
316	Julia	2022	10	9	6	12.4	-83.3	hurricane	1.0	75
317	Lisa	2022	11	2	21	17.4	-88.3	hurricane	1.0	80
318	Nicole	2022	11	9	23	26.6	-78.2	hurricane	1.0	65

[319 rows x 10 columns]

2eme dataset: MIN

	name	year	month	day	hour	lat	long	status	category	wind
0	Blanche	1975	7	27	6	35.9	-70.0	hurricane	1.0	65
1	Caroline	1975	8	30	0	23.3	-94.2	hurricane	1.0	65
2	Doris	1975	8	31	0	34.9	-46.3	hurricane	1.0	65
3	Eloise	1975	9	16	18	19.5	-68.4	hurricane	1.0	65
4	Faye	1975	9	26	0	26.5	-60.0	hurricane	1.0	65
..
314	Fiona	2022	9	18	12	17.2	-66.2	hurricane	1.0	65
315	Ian	2022	9	26	6	17.7	-81.7	hurricane	1.0	65
316	Julia	2022	10	9	0	12.5	-82.0	hurricane	1.0	65
317	Lisa	2022	11	3	0	17.5	-88.6	hurricane	1.0	65
318	Nicole	2022	11	9	23	26.6	-78.2	hurricane	1.0	65

[319 rows x 10 columns]

B. Le dataset des ouragans

Cependant, ce dataset manque d'une donnée très importante : la **température de la surface de l'eau**, qui est au cœur de notre questionnement.

Ainsi, nous avons dû trouver un dataset complémentaire avec les températures de surface de l'eau de l'océan Atlantique afin de pouvoir effectuer notre analyse.

Nous avons trouvé un site avec un fichier pour chaque année depuis 1981, avec les moyennes de température chaque jour de l'année selon la latitude et la longitude.

De ce fait, nous avons pu commencer nos analyses en créant des fonctions permettant de lier les données de ces deux datasets.

Notre première idée a été de télécharger tous les fichiers par année, puis afficher un graph représentant la moyenne de température de l'océan dans l'année. Cependant, au vu de la quantité de données, cela aurait pris beaucoup trop de temps, et n'aurait pas été pertinent en termes de complexité de chargement de données.

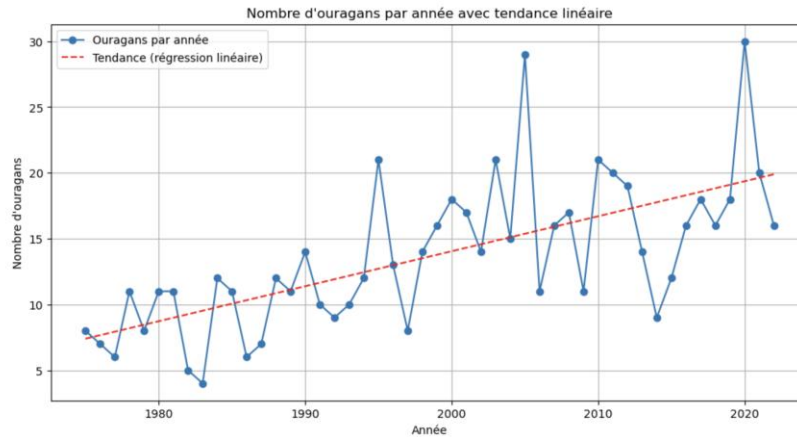
Notre deuxième idée a été de parcourir notre 1er dataset avec tous les ouragans, puis effectuer une recherche API sur le site en chargeant seulement les fichiers correspondant à l'année de l'ouragan étudié pour effectuer des analyses au cas par cas. Cependant, encore une fois, cela s'est avéré trop long et fastidieux.

Ainsi, nous avons décidé d'utiliser le graphique suivant, représentant le nombre d'ouragans par an sur la période 1975- 2022 afin de choisir quelques années clés de pics, ou à l'inverse, d'années où il n'y a pas eu beaucoup d'ouragans. La raison de cette sélection était de comparer les températures de surfaces de l'eau pendant cette période, et de montrer le lien entre la SST et la fréquence d'ouragans, ainsi que leur puissance.

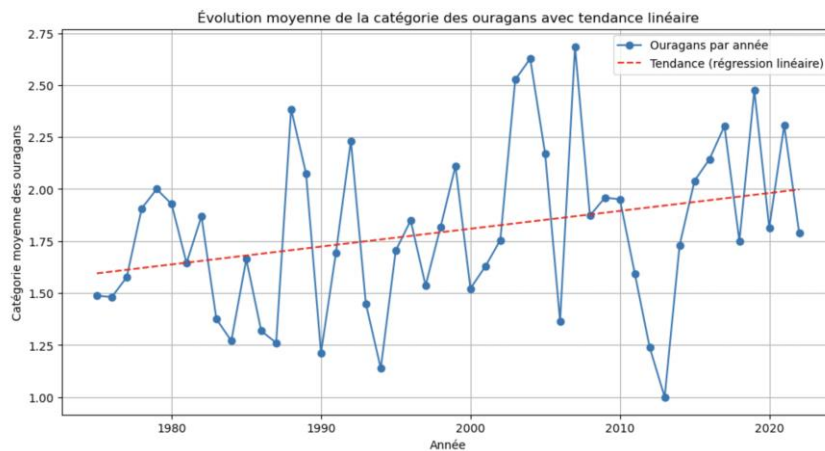
II. Utilisation des datasets et premières analyses

Nous avons commencé par générer des graphiques avec des tendances linéaires pour illustrer notre idée générale.

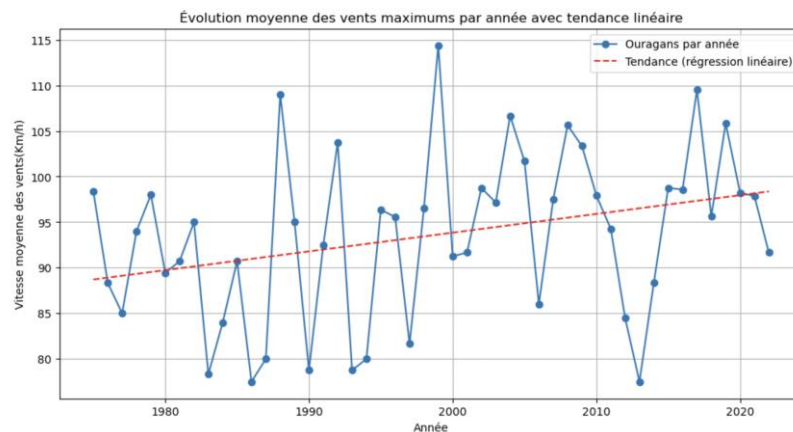
- Nous remarquons déjà une hausse de fréquence des ouragans au fil des années grâce à la régression linéaire en rouge.



- Nous remarquons également une hausse de la puissance des ouragans, cette tendance ressemble à celle de hausse de fréquence.



- Enfin, nous remarquons aussi une hausse de la vitesse des vents.



Nos analyses précédentes mettent en évidence un renforcement progressif des ouragans au fil des années, ainsi qu'une augmentation de leur fréquence.

Ensuite, nous avons créé une fonction modulaire [get_hurricane_year_by_index](#) qui nous permet

d'extraire l'année à laquelle a eu lieu un ouragan à partir de la ligne du fichier donnée en paramètre. Nous avons préféré présélectionner les dates renvoyées par cette fonction car il y aurait eu une trop grosse quantité de données à télécharger pour la suite de notre analyse, comme dit précédemment.

Nous avons également choisi, afin de gagner en visibilité, de faire une fonction

[plot_hurricane_on_map_with_category](#) permettant d'afficher sur une carte du monde la localisation exacte des ouragans, ainsi que leurs informations respectives (nom, date et catégorie).



Sur cette représentation, nous avons représenté la catégorie des ouragans grâce à un code couleur pour pouvoir mieux les identifier :

- Catégorie 1 : **bleu**
- Catégorie 2 : **vert**
- Catégorie 3 : **orange**
- Catégorie 4 : **rouge**
- Catégorie 5 : **noir**

Grâce à cette représentation, nous avons déjà pu identifier les hotspot d'ouragans. Ils se situent essentiellement dans le Golfe du Mexique et dans les Bermudes. Nous remarquons également que les ouragans les plus puissants sont essentiellement regroupés au nord de la Mer des Caraïbes en allant vers le Golfe du Mexique.

III. Analyse du lien entre nos datasets

La deuxième partie de notre analyse a consisté à parcourir l'ensemble de notre premier dataset (celui contenant tous les ouragans), et à charger dynamiquement, pour chaque ouragan, le fichier NetCDF

correspondant à son année. Bien que plus sélective, cette méthode demeurerait fastidieuse si l'on souhaitait comparer plusieurs années ou réaliser des analyses à grande échelle.

C'est pourquoi nous avons choisi une approche sélective :

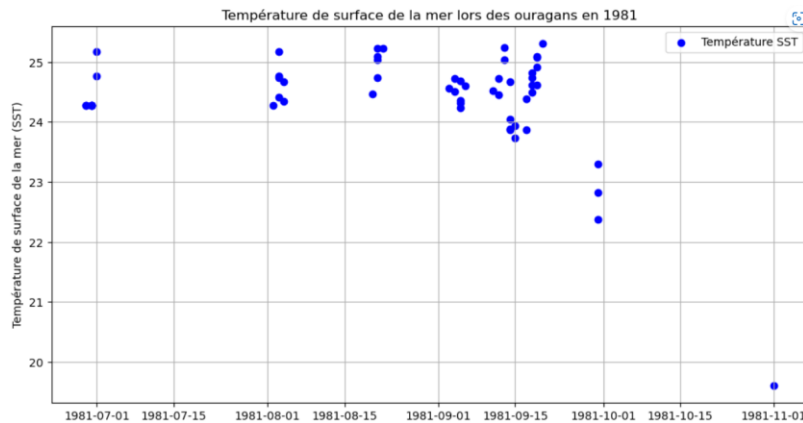
1. Nous avons d'abord observé la fréquence des ouragans par année (de 1975 à 2022) et identifié des années-clés : certaines avec un nombre élevé d'ouragans et d'autres avec un nombre plus faible.
2. Nous avons ensuite écrit la fonction `plot_sst_for_year(year)`, qui :
 - a. Filtre dans le CSV tous les événements se produisant l'année choisie.
 - b. Ouvre le fichier NetCDF (SST) pour cette année précise.
 - c. Extrait, pour chaque événement (ouragan), la température de surface de la mer (SST) au moment et au lieu exact de l'événement, en utilisant sa latitude et longitude.
 - d. Affiche un graphique **date (en abscisse) vs. température SST (en ordonnée)**, afin de visualiser la distribution des températures lors de ces ouragans.

Cette approche nous permet **d'éviter de charger tous les fichiers NetCDF** pour toutes les années, tout en mettant en évidence le lien entre la SST et l'activité des ouragans pour quelques années représentatives. Nous observons ainsi que, durant les périodes où la SST est plus élevée (souvent au-delà de 26,5 °C), les ouragans sont plus nombreux ou plus intenses, ce qui conforte l'hypothèse que la température de surface de la mer est un facteur crucial dans la formation et le développement des ouragans.

Grâce à ce procédé, **nous parvenons à comparer, de manière ciblée**, les années marquantes (pics d'ouragans) avec celles moins actives, tout en limitant la quantité de données chargées et en simplifiant la logistique de traitement.

Voici un **exemple illustratif** pour l'année 1981, où nous avons tracé la température de surface de la mer (SST) lors de chaque événement d'ouragan. Nous avons **répété la même analyse pour d'autres années** clés afin de comparer la SST et l'activité cyclonique sur différentes périodes.

Ce graphique illustre la **température de surface de la mer (SST) lors des ouragans en 1981**. Plus précisément :



1. **Axe des abscisses (x)** : Il représente les dates auxquelles les ouragans ont été observés. On voit ici que la plupart des événements se concentrent entre le début de l’été et la fin de l’automne 1981.
2. **Axe des ordonnées (y)** : Il indique la température de surface de la mer (SST) en degrés Celsius. Les points se situent globalement entre 20 °C et 30 °C, ce qui correspond aux températures de surface typiques dans les zones où se forment les ouragans.
3. **Points bleus** : Chaque point correspond à un événement d’ouragan (issu du fichier CSV) pour lequel la fonction a extrait la température la plus proche (en temps et en espace) depuis le fichier NetCDF. Ainsi, on visualise directement la SST au moment et à l’endroit où s’est produit l’ouragan.
4. **Interprétation** : On constate que la majorité des ouragans (points) se produisent lorsque la SST dépasse généralement 26–27 °C, ce qui est cohérent avec le seuil critique souvent cité (environ 26,5 °C) pour favoriser le développement des cyclones tropicaux. Les dates plus avancées dans la saison (octobre, voire début novembre) montrent parfois des SST plus basses, reflétant le refroidissement progressif de la surface de l’océan.

En résumé, ce **nuage de points** permet de repérer, pour chaque événement de tempête, la **température de l’eau** correspondante, mettant en évidence le rôle joué par une SST suffisamment élevée dans la formation et l’intensification des ouragans.

III. Études approfondies

Nous souhaitons maintenant, faire des analyses plus poussées :

1- Trouver l'ouragan le plus fort, sa durée, le mois pendant lequel il est survenu

Nous avons défini le critère de force de l’ouragan en analysant les colonnes “wind” et “pressure” de notre dataset. Après filtrage, il en ressort que l’ouragan le plus fort (de 1975 à 2022) est l’ouragan “Allen”

survenu en Août 1980 et actif pendant 4 jours. Sa vitesse de vent était de 165 Km/h et la pression de 945 hPa

2- Trouver l'ouragan le moins fort, sa durée, le mois pendant lequel il est survenu

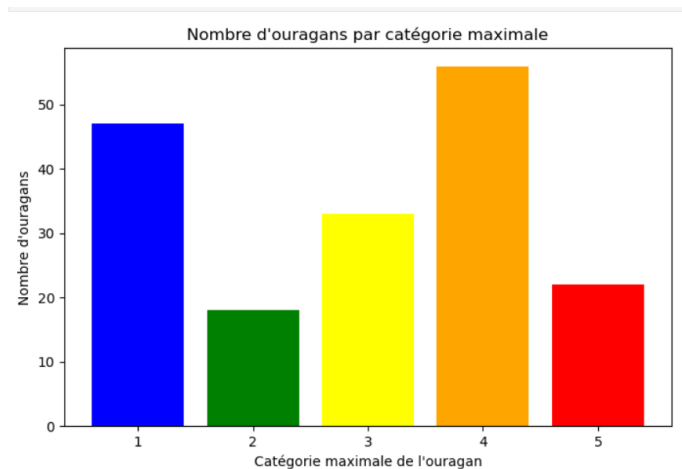
L'on fait de même pour trouver l'ouragan le moins fort. Il en ressort que l'ouragan "*Debby*" est l'ouragan le moins fort ; survenu le mois d'Août de l'an 2000, il a été actif pendant seulement 1 jour.

3- Généraliser cela en montrant que les ouragans ont tendance à se produire à une certaine période de l'année et qu'il est rare qu'ils se produisent à une autre certaine période de l'année

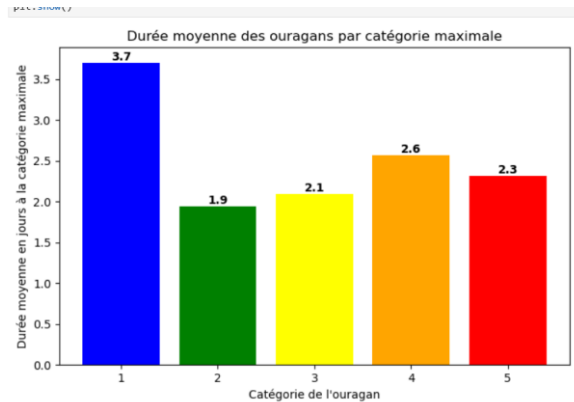
3-1) Quelle est la durée moyenne des ouragans ?

Vu que pour un ouragan, on peut avoir plusieurs lignes (donc plusieurs enregistrements) dans notre dataset, nous nous sommes rendu compte que tous les ouragans de notre dataset en termes de puissance, avaient commencé à la catégorie 1. Alors nous considérons comme ouragans de catégorie 5, les ouragans qui ont atteint au moins une fois la catégorie 5 pendant leur durée. Et ce, respectivement pour les ouragans de catégorie 4, 3 et 2. On ne garde que la catégorie maximale que l'ouragan a atteint tout le long de sa durée.

Ainsi, nous obtenons le nombre **exact** d'ouragans par catégorie !

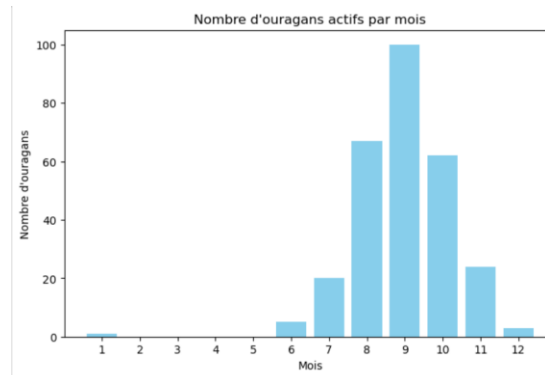


Puis par de manipulations habiles, nous obtenons le graphe ci-dessous, qui représente la durée moyenne des ouragans par catégorie maximale, sur lequel nous pouvons remarquer que les ouragans les plus faibles (de catégorie 1) sont ceux qui ont une durée d'activité la plus élevée.



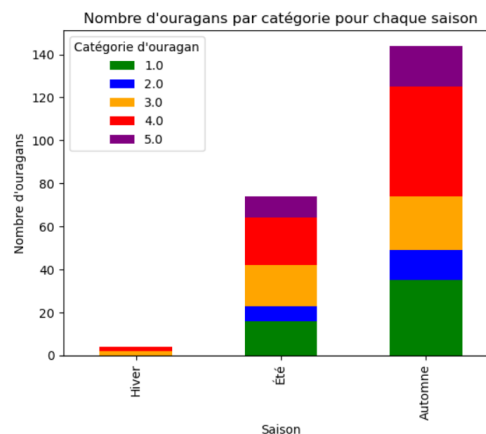
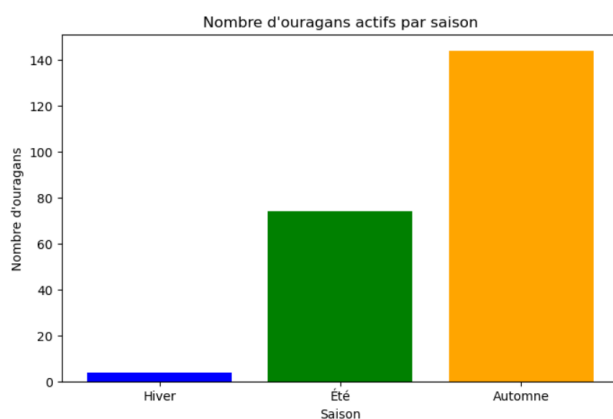
3-2) Quelle est la période de l'année la plus propice aux ouragans ?

Pour répondre à cette question, on commence par regarder pour chaque ouragan, quels sont leur mois d'apparition.



On remarque que les mois d'activités des ouragans sont 'Janvier', puis de 'Juin' à 'Décembre' avec un pic d'activité en 'Septembre'. En revanche de 'Février' à 'Mai', on observe aucune activité d'ouragan.

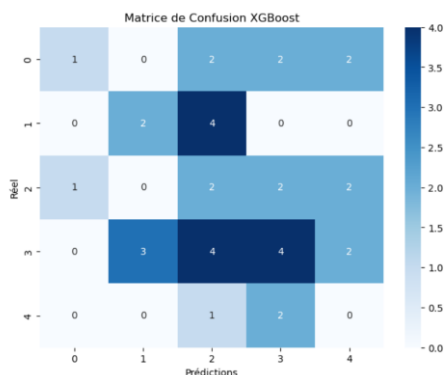
L'on décide de faire un graphique saisonnier de l'apparition des ouragans (voir graphiques ci-dessous)



On remarque que les ouragans apparaissent beaucoup plus en 'Automne', moins en 'Été', encore moins en 'Hiver', et jamais au 'Printemps' (Mars, Avril, Mai).

3-3) Prédire la catégorie maximale que peut atteindre un ouragan lors de son activité

Nous avons créé un modèle de prédiction 'XGBClassifier' qui est capable de prédire la catégorie maximale qu'un ouragan peut atteindre compte tenu de sa durée dans le temps, son mois et sa saison d'apparition. Il prédit 21% de bonnes catégories. Ci-dessous la matrice de confusion...



Si améliorées, les prédictions de notre modèle pourraient être utiles dans la prédiction de la catégorie maximale qu'un ouragan peut atteindre compte tenu de sa durée, du mois et de la saison de son apparition. Cela permettrait de mieux anticiper son intensité et préparer des évacuations massives si nécessaire

Sources

- [1] Chan, Duo, et al. ["Improved simulation of 19th-and 20th-century North Atlantic hurricane frequency after correcting historical sea surface temperatures."](#) Science advances (2021)
- [2] Dataset des tempêtes de l'Atlantique Nord, [NOAA's storm dataset](#)
- [3] Différence entre un typhon, un ouragan et un cyclone, [National Geographic](#)
- [4] Ouragans : Définition, formation et catégories, [PojetEcolo](#)
- [5] Dataset des SST (Sea Surface Temperature), [NOAA's temperature dataset](#)

