

Dossier Techniques de Prévision et Conjoncture  
MASTER 1 ECAP

**Prévision de l'indice du chiffre d'affaire dans les services  
touristiques à partir des données récoltés entre août 2018 et  
novembre 2023**

Dossier réalisé par :  
Orhan Cassandre  
Trillaud Valorys

Année universitaire 2023-2024

# Sommaire

1 Introduction	2
2 Analyse exploratoire	3
3 Estimation et prévision	11
4 Comparaisons des modèles	17
5 Le meilleur modèle	25
6 Conclusion et discussion	28
7 Annexe	29
8 Bibliographie	35
9 Table des matières	36

# 1 Introduction

Le secteur du tourisme en France est un pilier majeur de l'économie, générant "pas moins de 2 millions d'emplois directs et indirects"<sup>1</sup>. Il est donc indéniable que le tourisme occupe une place prépondérante dans le tissu économique français. En outre, ses recettes, qui "se sont élevées à 63,5 milliards d'euros"<sup>2</sup> en 2023, ce secteur "représente environ 7,5 % du PIB français"<sup>3</sup>. Ces chiffres imposants soulignent l'importance du tourisme dans l'économie française. Le tourisme ne se limite pas à l'accueil des touristes étrangers, mais englobe également les touristes intérieurs, les loisirs et les activités culturelles. Des millions de visiteurs vont chaque année visiter des lieux emblématiques de la France tels que la Tour Eiffel, le Louvre ou encore la Côte d'Azur.

Prédire les tendances de ce secteur peut alors permettre de comprendre l'évolution d'un secteur majeur en France. De plus, le tourisme est un vecteur puissant de rayonnement international pour la France. En mettant en valeur son histoire, sa cuisine, son art et son patrimoine, le tourisme renforce l'image du pays à l'étranger, attirant des investissements et soutenant la diplomatie culturelle. Ainsi, notre analyse vise à mieux comprendre et à anticiper les tendances de ce secteur dynamique. Nous allons alors chercher à trouver la meilleure prévision de l'index du chiffre d'affaires dans les services.

Pour ce faire, nous débuterons par une analyse statistique approfondie de nos données, permettant d'extraire les informations clés et de corriger la série si besoin en présence de points atypiques. Nous examinerons ensuite la saisonnalité de cette série avant de passer à l'estimation des modèles et la prévision des données à travers différentes méthodes. Cette démarche nous permettra de comparer les méthodes selon différents indicateurs, afin d'obtenir la meilleure qualité de prévision et précision.

---

1 <https://www.entreprises.gouv.fr/fr/tourisme> (consulté le 24/04/2024)

2 <https://www.atout-france.fr/fr/actualites/note-de-conjoncture-mars-2024> (consulté le 24/04/2024)

3. <https://www.campusfrance.org/fr/actu/tourisme-2023-une-annee-record-pour-la-france#:~:text=Car%2C%20rappelle%20le%20site%20pr%C3%A9sidentiel,%2C5%25%20du%20PIB%20fran%C3%A7ais>. (consulté le 24/04/2024)

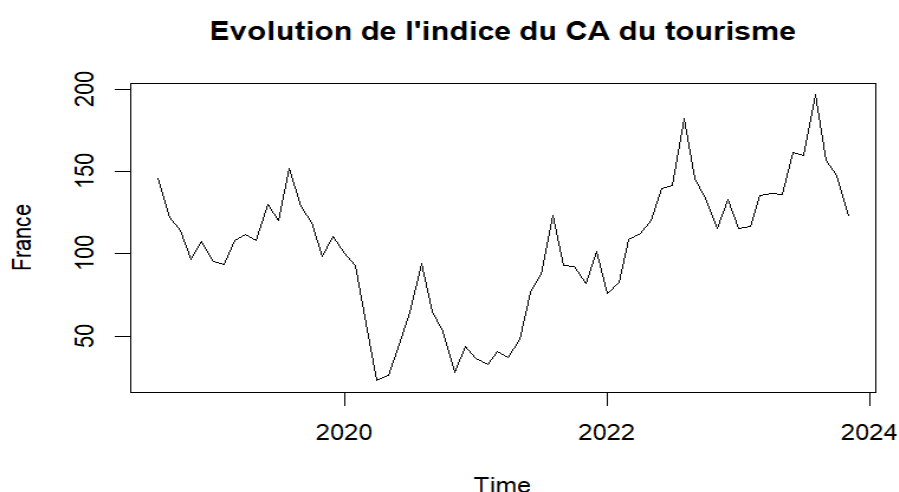
## 2 Analyse exploratoire

### 2.1 Présentation de la série

Dans cette analyse, notre objectif est de prévoir l'évolution de l'indice du chiffre d'affaires dans les services touristiques. Les données utilisées pour cette prévision proviennent d'Eurostat<sup>4</sup>, l'Office statistique de l'Union européenne, qui collecte et publie des statistiques économiques et sociales pour les pays membres de l'Union européenne. Dans notre étude, les services englobent plusieurs secteurs d'activités : le transport aérien, l'hébergement, les agences de voyages, les tours opérateurs et les autres services de réservation et activités connexes.

L'indice du chiffre d'affaires dans les services touristiques mesure la performance financière globale des services liés au tourisme. Pour cette analyse, nous utiliserons des données mensuelles. Cet indice est calculé en comparant les recettes des entreprises actives dans ces secteurs sur une période donnée par rapport à une période de référence, où la base 100 correspond à l'année 2015.

*Figure n°1 : Graphique d'évolution de la série*



En examinant la figure n°1, nous constatons une chute importante du secteur du tourisme, attribuable aux différents confinements liés à la pandémie du Covid-19.

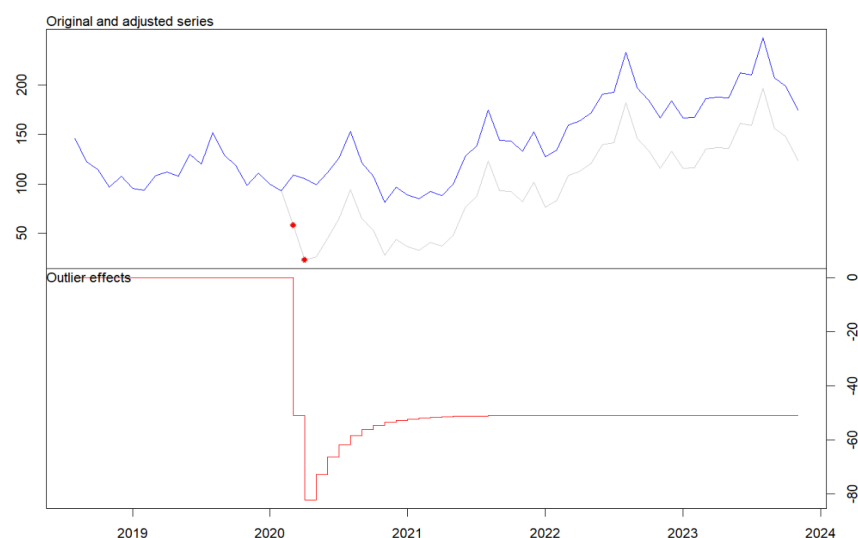
<sup>4</sup> [https://ec.europa.eu/eurostat/databrowser/view/sts\\_setu\\_m\\$dy\\_667/default/table?lang=en&category=tour.tour\\_inda.tour\\_sts](https://ec.europa.eu/eurostat/databrowser/view/sts_setu_m$dy_667/default/table?lang=en&category=tour.tour_inda.tour_sts) (consulté le 24/04/2024)

En août 2020, nous observons un pic lié à l'assouplissement des mesures de confinement, mais le tourisme connaît ensuite un nouveau déclin avant de reprendre une croissance à partir de mi-2021. Cependant, avant et après la période du Covid-19, le secteur présente des fluctuations tout au long de l'année. Il est intéressant de noter que les mois de juillet et août correspondent aux périodes où le secteur du tourisme est le plus dynamique. Enfin, depuis 2021, nous pouvons remarquer une tendance à la hausse du secteur du tourisme.

## 2.2 Détection des points atypiques

La détection des points atypiques dans une série saisonnière est une étape cruciale dans l'analyse des données temporelles en raison de l'impact que ses points peuvent avoir sur les estimations et les modèles statistiques. Les points atypiques, qu'ils soient transitoires ou permanents, peuvent engendrer divers problèmes d'estimation pour les méthodes de moindres carrés ordinaires (MCO) ou encore les modèles ARIMA. La détection et la correction appropriées des points atypiques dans les séries saisonnières sont essentielles pour assurer la robustesse des analyses et des prévisions, c'est pour cela que nous débutons notre analyse par cette étape cruciale.

Figure n°2 : Représentation graphique des points atypiques



*Figure n°3 : Tableau descriptif des points atypiques*

Outlier	Type	Date	Coefhat	tstat
1	LS	Mars 2020	-50,98	-7,408
2	TC	Avril 2020	-31,16	-4,996

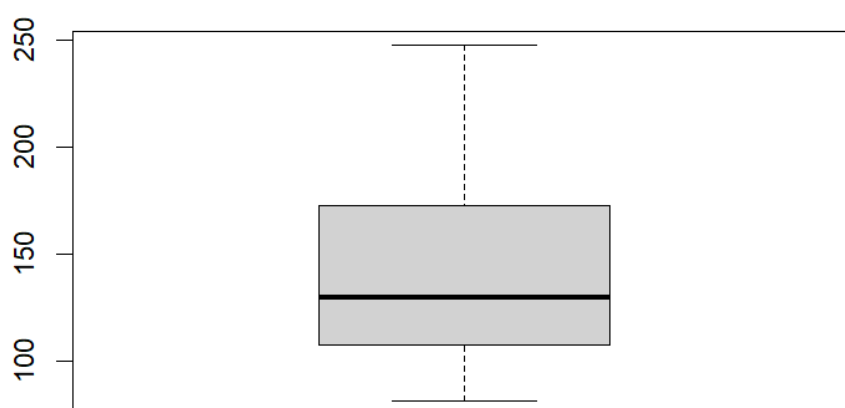
La figure n°2 et le tableau nous permettent d'identifier deux points atypiques. Le premier correspond à un changement de niveau (LS: level shift) en mars 2020. Cela signifie que nous avons un changement de niveau abrupt et permanent sans retour au niveau initial. Ce changement est attribuable au premier confinement qui a débuté le 17 mars 2020. Le deuxième point atypique, repéré en avril 2020, correspond à un changement transitoire (TC: transitory change). Un TC indique un changement de niveau suivi d'un retour progressif à la valeur initiale. Ce point atypique est dû au confinement généralisé en France tout le long du mois d'avril de 2020, entraînant l'arrêt du secteur du tourisme. Cependant, dès le 11 mai, les activités ont pu reprendre progressivement, y compris dans le domaine touristique.

La détection de points atypiques permet également d'ajuster la série en corrigeant la présence de points atypiques. Par conséquent, la suite de notre analyse portera sur cette série corrigée, représentée en bleu sur la figure n°2. Cette visualisation met en évidence la distinction entre la série originale et la série corrigée. En effet, la série corrigée ne présente pas de chute en mars 2020, contrairement à la série d'origine.

## 2.3 Statistiques descriptives

Tout d'abord, examinons le boxplot de la série corrigée. Nous constatons que celui-ci ne présente aucun point atypique.

*Figure n°4 : Boxplot de la série corrigée*



En outre, l'analyse des statistiques nous offre une vue d'ensemble de notre série. En observant le tableau des statistiques descriptives, nous notons que la médiane et la moyenne présentent une différence notable, suggérant une possible asymétrie de la distribution. Pour quantifier cette asymétrie, nous examinons la skewness et la kurtosis. La kurtosis négative (-0,705) suggère une distribution plus aplatie que celle d'une distribution normale, tandis qu'une skewness positive (0,558) indique une asymétrie vers la droite. Ces valeurs indiquent donc une légère asymétrie dans notre série. Les quartiles (Q1 = 107,83 et Q3 = 172,21) confirment cette observation, montrant une dispersion importante des données autour de la médiane. De plus, l'écart-type (41,22) et la variance (1 698,9) sont très élevés, ce qui témoigne d'une dispersion importante des valeurs, s'étalant de 81,35 à 247,58 pour une base 100 en 2015.

*Figure n°5 : Tableau des statistiques descriptives*

Min	Q1	Médiane	Moyenne	Q3	Max	Kurtosis	Skewness	Ecart-type	Variance
81,35	107,83	129,50	140,35	172,21	247,58	- 0,705	0,558	41,22	1 698,9

## 2.4 Détection de la saisonnalité

Afin de vérifier la présence de saisonnalité dans notre série mensuelle corrigée, nous utilisons plusieurs tests statistiques spécialisés. Deux de ces tests sont particulièrement pertinents : le test de Webel-Ollech et le F-test sur les variables dichotomiques saisonnières (Seasonal Dummies). Dans ces tests, l'hypothèse nulle correspond à l'absence de saisonnalité.

Le test de Webel-Ollech est une méthode robuste qui combine deux tests distincts pour évaluer la saisonnalité. Tout d'abord, le test QS modifié examine si la série présente des autocorrélations positives significatives aux retards saisonniers, indiquant une répétition périodique des motifs. Ensuite, le test de Kruskal-Wallis analyse les différences significatives entre les rangs moyens à des périodes spécifiques pour les observations, ce qui permet de détecter des modèles saisonniers dans les données. En rejetant l'hypothèse nulle d'absence de saisonnalité, ces tests confirment que notre série mensuelle corrigée présente bel et bien une saisonnalité.

En complément, le F-test sur les variables dichotomiques saisonnières est une approche classique pour évaluer la saisonnalité. Il examine si l'ajout de variables binaires pour chaque saison améliore significativement le modèle. Une p-value inférieure à 0,05 indique que l'hypothèse nulle d'absence de saisonnalité est rejetée, corroborant ainsi la présence de saisonnalité dans notre série.

*Figure n°5 : Test de la saisonnalité*

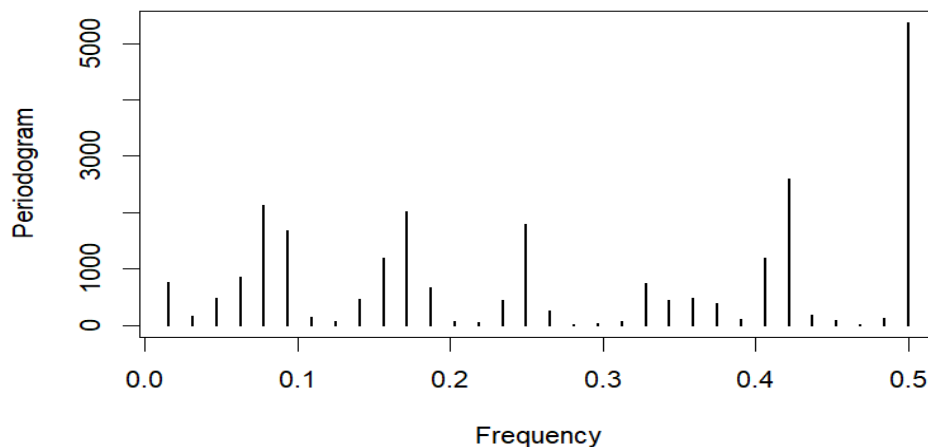
Test	P-value
Seasonal Dummies	0
Webel-Ollech	1,11e-16

De plus, nous utilisons la représentation du périodogramme sur la série en différences premières pour confirmer graphiquement la saisonnalité. Les pics significatifs dans le périodogramme révèlent des fréquences périodiques, mettant en



lumière la répétition des motifs dans les données et confirmant visuellement la présence de saisonnalité.

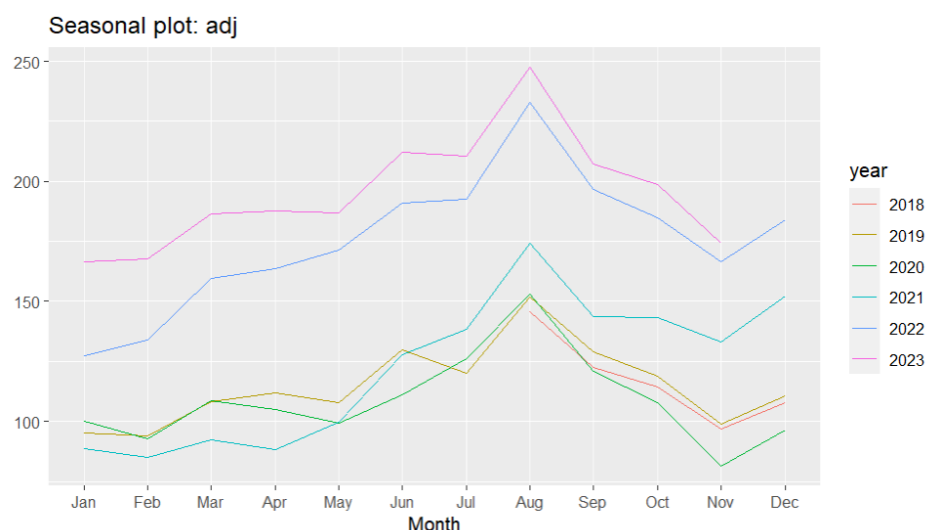
Figure n°6 : Périodogramme sur la série en différence première



En combinant ces approches statistiques et graphiques, nous pouvons affirmer de manière robuste que notre série mensuelle corrigée présente une saisonnalité significative.

Maintenant que nous avons confirmé la présence de saisonnalité dans notre série, nous cherchons à déterminer si son modèle de décomposition est additif ou multiplicatif. Pour ce faire, nous adoptons la méthode du profil, qui nous permet de comparer les différentes années entre elles. Cette approche nous conduit à identifier un modèle de décomposition de type additif. Cette conclusion est étayée par l'observation des courbes annuelles qui sont parallèles entre elles.

Figure n°7 : Graphique des courbes annuelles



Pour consolider notre choix, nous utilisons le test log-level, qui évalue la relation entre le niveau de la série et la saisonnalité. Ce test nous permet de confirmer la nature additive du modèle de décomposition. En effet, l'absence de transformation logarithmique confirmée par l'aicdiff inférieur à 0 implique que le modèle n'est pas multiplicatif mais bien additif.

*Figure n°8 : Test log-level*

<b>tfunction</b>	<b>adjust</b>	<b>aicdiff</b>
Auto	None	-2

En utilisant ces deux méthodes, nous sommes en mesure de déterminer de manière robuste que le modèle additif est le plus approprié pour la décomposition de notre série saisonnière.

## 2.5 Désaisonnalisation et décomposition

À partir de maintenant, nous allons faire notre analyse sur la série ajustée à laquelle nous avons retiré les 12 derniers mois. Cette démarche nous permettra de comparer les prévisions aux valeurs observées de cette même série ajustée.

Maintenant que nous savons que la série suit un modèle additif, nous pouvons désaisonnaliser et décomposer la série grâce à la méthode X13-ARIMA-SEAT.

En quelques mots, la désaisonnalisation va nous permettre d'identifier et de supprimer la saisonnalité de notre modèle afin de mettre en évidence les autres éléments, comme les tendances ou les cycles, qui jouent un rôle important dans l'analyse économique. La méthode que nous allons utiliser pour désaisonnaliser notre série se base sur une méthode semi-paramétrique basée sur des modèles hybrides, combinant à la fois des modèles implicites, fondées sur des filtres de lissage linéaire, et des modèles explicites, ou les composantes adoptent soit un comportement déterministe ou purement linéaire.

La méthode que nous allons utiliser est X13-ARIMA-SEATS, avant d'utiliser cette méthode, nous devons effectuer un prétraitement de nos données en utilisant la régression ARIMA qui va nous permettre de rendre la série stationnaire, de supprimer les effets déterministes, outliers, etc et finalement nous permettre par la suite l'estimation, la prévision et l'interpolation de modèles de régression.

Nous avons constaté que notre modèle ne prend pas en compte les effets déterministes dans l'ajustement de la série. En effet, notre modèle ne comporte pas de terme moyen, aucun effet des jours ouvrables, des années bissextiles et de Pâques, de plus aucun point aberrant n'est détecté. Le module de prétraitement nous a ainsi donné un modèle SARIMA(0,1,0)(0,1,1), ce qui signifie que nous avons une différenciation non-saisonnière d'ordre 1 et une différenciation saisonnière d'ordre 1.

## 3 Estimation et prévision

### 3.1 Méthode de prévision sans lissage exponentiel

Pour estimer et prévoir nos modèles, nous allons utiliser des méthodes de désaisonnalisation des séries temporelles. Dans cette section, nous nous concentrerons spécifiquement sur la méthode de désaisonnalisation sans lissage exponentiel.

#### 3.1.1 Méthode naïve et naïve saisonnière

La méthode naïve est une approche simple de la prévision des séries temporelles, reposant sur l'hypothèse que toutes les observations futures seront égales à la dernière observation connue. Cependant, cette simplicité comporte des limites, elle ne prend pas en compte les tendances, les saisons ou les cycles présents dans la série. Ainsi, dans notre contexte, où notre série temporelle présente une saisonnalité, l'application de cette méthode risque de conduire à des prévisions inexactes. En négligeant ces composantes clés des données, la méthode naïve peut sous-estimer ou surestimer les valeurs futures de notre série. C'est pourquoi nous nous sommes dirigés vers la méthode naïve saisonnière qui prend en compte la saisonnalité.

#### 3.1.2 Méthode non paramétrique : STL

La méthode STL (Seasonal-Trend decomposition based on Loess) est une approche non paramétrique pour décomposer une série temporelle en ses composantes saisonnières, tendanciennes et résiduelles. Fondée sur une technique de régression locale, elle utilise la méthode des  $k$  plus proches voisins pour estimer avec précision la tendance et la saisonnalité de la série.

Le modèle STL peut être enrichi en combinant la décomposition saisonnière et de tendance STL avec le modèle ETS, offrant ainsi une approche encore plus robuste pour la modélisation et la prévision des séries temporelles. Dans notre cas, le modèle estimé suit le format STL + ETS(A,N,N), ce qui indique que notre modèle

ne contient ni tendance ni de saisonnalité, tandis que les erreurs du modèle sont additives. Le paramètre alpha est estimé à 0,99, indiquant un lissage fort, où les valeurs passées ont un poids élevé dans la prédiction de la tendance future. L'écart-type de l'erreur du modèle est de 6,1226, ce qui indique une variabilité plutôt importante des données autour de la tendance et de la saisonnalité prédite par le modèle.

### 3.1.3 Méthode semi-paramétrique : X13-ARIMA-SEATS

La méthode X13-ARIMA-SEATS est une approche puissante pour modéliser les séries temporelles en utilisant un processus SARIMA (Seasonal Autoregressive Integrated Moving Average) et en ajustant les données en début et en fin de série. Cette méthode est particulièrement efficace pour les séries temporelles stationnaires une fois que les effets déterministes ont été supprimés. De plus, le module RegARIMA impose des contraintes strictes sur les paramètres du modèle SARIMA(p,d,q)(P,D,Q)s, à savoir :  $p, q \leq 3$ ,  $P, D \leq 1$ ,  $d \leq 3$  et  $D \leq 2$ .

Le modèle estimé par la méthode X13-ARIMA-SEATS n'a pas subi de transformation logarithmique, ce qui signifie que la décomposition du modèle est additive. De plus, le modèle de régression n'a aucun effet de moyenne, d'effet des jours de trading, d'effet des années bissextiles ou d'effet de Pâques, et aucune valeur aberrante n'a été incluse dans le modèle de régression. Il s'agit d'un modèle avec une régression et un modèle ARIMA(0,1,0)(0,1,1). Le coefficient du Bêta saisonnier est estimé à -0,8695, indiquant une tendance à la baisse dans le temps. La p-value très faible indique que ce coefficient est statistiquement significatif au seuil de 1%. Nous avons une différenciation première et une différenciation première saisonnière.

### 3.1.4 Méthode paramétrique : STS

Une des méthodes paramétriques pouvant être utilisées est l'approche STS (Séries Temporelles Structurelles). Cette approche permet de spécifier directement des modèles ARIMA pour chaque variable inobservable. La décomposition se fait

alors en deux étapes distinctes : l'équation d'observation et le modèle structurel sont mis sous forme d'espace d'état.

Nous pouvons alors interpréter 4 des paramètres de notre modèle estimés par la méthode STS. Tout d'abord le level, il indique la variabilité de la composante de niveau dans la série temporelle. Une variance élevée de 96,423, suggère que les valeurs observées fluctuent considérablement autour de la tendance centrale de la série. La variance de la pente (slope) est relativement élevée, avec une valeur de 0,466. Cela suggère une variabilité accrue dans les changements de direction ou de pente de la série temporelle. Pour la variance de la saisonnalité qui vaut 4,564, cela indique une variabilité saisonnière assez prononcée de la série, avec des variations significatives selon les périodes de l'année. Enfin, la variance de l'épsilon étant de 0, cela suggère que la variabilité résiduelle est nulle, les autres composantes du modèle expliquent donc efficacement la variation des données.

### 3.1.5 Bagged Model

Le bagged model, également connu sous le nom de modèle ensembliste par agrégation bootstrap. Il exploite d'abord le bootstrapping pour générer plusieurs échantillons représentatifs à partir des séries temporelles. Ensuite, il applique la transformation de Box-Cox pour stabiliser la variance et la décomposition STL pour extraire les composantes saisonnières, tendanciennes et résiduelles des données. Cette approche permet de traiter efficacement les séries temporelles, en prenant en compte à la fois les variations saisonnières et les tendances grâce à une méthode de lissage ETS .

### 3.1.6 SARIMA (p,q,d)(P,D,Q)<sub>12</sub>

Le modèle SARIMA(p,d,q)(P,D,Q)<sub>s</sub> est une extension du modèle ARIMA qui incorpore à la fois les composantes non-saisonnières et saisonnières de la série temporelle. La composante saisonnière "s" représente la périodicité des données ; par exemple, pour des données mensuelles, s est égale à 12. SARIMA combine les composantes autoregressives (AR), intégrées (I) et à moyenne mobile (MA) du modèle ARIMA avec des termes saisonniers pour modéliser les motifs saisonniers récurrents dans les séries chronologiques.

Dans notre analyse, nous avons obtenu un modèle SARIMA(0,1,0)(0,1,1)<sub>12</sub>. Le choix de  $s=12$  est approprié, car nos données sont mensuelles. Les valeurs de  $d$  et  $D$ , toutes deux égales à 1, indiquent une modification de la série par une différenciation première et une différenciation saisonnière. Enfin, le seul paramètre significatif est  $Q=1$ , ce qui indique la présence d'une seule moyenne mobile saisonnière. Le coefficient associé à cette composante est de -0,87, ce qui indique que la composante est négativement corrélée avec les données.

## 3.2 Méthode de prévision avec lissage exponentiel

Nous allons maintenant utiliser la méthode de lissage exponentiel pour la prévision de nos données temporelles. Cette méthode permet de lisser les séries chronologiques en accordant différents niveaux d'importance aux valeurs passées.

### 3.2.1 Holt-Winters

Le lissage exponentiel est une méthode de lissage permettant la prévision à court terme accordant plus ou moins de poids aux valeurs du passé. L'estimation du lissage exponentiel double, utilisant la méthode de Holt Winter saisonnière additive. Révèle dans notre cas, des coefficients de lissage :  $\alpha$ ,  $\beta$  sont compris entre 0 et 1, la condition de ce modèle est donc bien respectée pour ces paramètres, mais nous obtenons un gamma égale à 1. Un alpha de 0,67 signifie que le modèle estimé tend vers une prévision rigide, l'influence des observations passées est d'autant plus importante et remonte loin dans le passé. Un bêta de 0,11 indique un lissage plus faible de la tendance, permettant aux prévisions de fluctuer davantage. Nous obtenons un gamma égal à 1, le lissage de la composante saisonnière est très fort. Cela signifie que la variation saisonnière est entièrement capturée par le modèle et qu'il réagit très rapidement aux changements de cette composante saisonnière. Cette observation est logique dans notre cas, car les fluctuations saisonnières dans l'industrie du tourisme peuvent être fortement influencées par des facteurs tels que les vacances scolaires, les saisons touristiques traditionnelles et les conditions météorologiques, ce qui peut entraîner des modèles de saisonnalité très prononcés et prévisibles.

### 3.2.2 ETS

La méthode ETS (Error-Trend-Seasonal) ajoute un terme d'erreurs à l'espace d'état de la méthode saisonnière de Holt-Winters. L'estimation de notre modèle indique un ETS(M,N,A), ce qui signifie que le lissage exponentiel simple donne une saisonnalité additive et des erreurs multiplicatives sans tendance. Les différents paramètres estimés par notre modèle,  $\alpha$ , utilisé pour modéliser la composante de niveau et  $\gamma$ , pour modéliser la saisonnalité, respecte la condition  $0 < \alpha, \gamma < 1$ .

### 3.2.3 TBATS

La méthode TBAT, une extension des modèles ETS, offre une solution robuste pour traiter des séries temporelles avec des caractéristiques non-linéaires, une autocorrélation des résidus et divers comportements saisonniers. Introduisant une représentation trigonométrique des composantes saisonnières, le modèle TBATS( $\alpha$ , (p, q),  $\phi$ , {m1, k1}, {m2, k2}, . . . , {mT, kT }) propose une approche flexible pour la modélisation. Dans notre cas, avec des paramètres spécifiques, notamment un coefficient  $\lambda$  de 0,262 suggérant une légère transformation logarithmique pour stabiliser la variance, des ordres AR et MA nuls, indiquant l'absence de termes autorégressifs et de moyenne mobile, et un coefficient  $\phi$  de 0,854, dénotant une dépendance linéaire significative entre les observations passées et actuelles. De plus, la notation <12,5> spécifie une saisonnalité sur une période de 12 mois.

### 3.2.4 ADAM ETS

Le modèle ADAM ETS, également une extension du modèle ETS, possède l'avantage d'utiliser un vecteur de décalages dans la forme d'espace d'état, ce qui permet de réduire les dimensions des matrices et de rendre le modèle plus parcimonieux. De plus, ce modèle permet d'utiliser des distributions différentes de la loi normale pour le terme d'erreur, telles que la loi Laplace, log-normale ou gamma.

Cette méthode ADAM ETS permet d'estimer un modèle ETS(A,N,A) à partir d'une loi normale, cela signifie que les erreurs et la saisonnalité du modèle sont additives. Les différents paramètres estimés par notre modèle, ne respectent pas la



condition  $0 < \alpha$ ,  $\gamma^* < 1$ , puisque  $\alpha = 0$  et  $\gamma = 1$ , cela peut conduire à des problèmes au niveau de l'interprétation des résultats et présente donc un biais dans notre analyse.

### 3.2.5 ADAM ETS SARIMA

Nous pouvons également intégrer la méthode SARIMA au modèle ADAM ETS. Cette combinaison, connue sous le nom de méthode ADAM ETS SARIMA, permet d'estimer un modèle avec une fonction ETS et SARIMA.

Nous obtenons ainsi un modèle  $ETS(M,A,N) + SARIMA(1,0,0)[12]$ , ce qui signifie qu'il y a un lissage avec des erreurs multiplicatives et une tendance additive. De plus, le modèle contient un paramètre autorégressif saisonnier d'ordre 1.

Les différents paramètres estimés par notre modèle respectent les conditions puisque  $\phi = 0,886$  donc  $0,8 < \phi < 0,98$  et  $\alpha = 0,9997$  donc  $0 < \alpha < 1$ .

### 3.2.6 SSARIMA

Le modèle SSARIMA, ou modèle ARIMA d'espace d'état est reconnu pour sa performance accrue sur de petits échantillons et sa flexibilité.

Dans notre analyse, nous obtenons un modèle  $SARIMA(0,1,0)[1](0,0,2)[12]$ , cela indique que nous avons une différenciation première mais pas de différenciation saisonnière. De plus, aucun paramètre n'est présent dans la partie classique du modèle, tandis que nous avons deux paramètres MA dans la partie saisonnière. Les deux coefficients SMA sont positifs indiquant une corrélation positive entre ces deux composantes et les données.

### 3.2.7 CES

La méthode CES correspond à un modèle de lissage exponentiel complexe, basé sur la théorie des variables complexes. Cela permet de modéliser des séries temporelles, qu'elles soient stationnaires ou non.

Notre modèle inclut un type de saisonnalité "partial", ce qui signifie une saisonnalité partielle avec des composantes saisonnières réelles, équivalentes à la saisonnalité additive.

## 4 Comparaisons des modèles

### 4.1 Meilleur modèle estimé

Nous avons estimé plusieurs modèles afin de réaliser des prévisions sur une période d'un an. Le choix du modèle le plus approprié revêt une importance cruciale dans cette démarche. À cette fin, nous avons recours au critère AIC et à son homologue corrigé, l'AICc (AIC corrigé). Ces deux métriques nous permettent d'évaluer la qualité d'ajustement d'un modèle tout en tenant compte de sa complexité. Opter pour le modèle présentant le plus faible AIC et AICc nous permet de concilier un ajustement optimal tout en minimisant la complexité. En définitive, l'utilisation de l'AIC et de l'AICc dans le processus de sélection de modèles vise à garantir une modélisation statistique robuste et efficace pour nos besoins de prévision.

Afin de permettre la comparaison des méthodes utilisées, il a été nécessaire pour certaines méthodes de se reposer sur les formules de l'AIC et de l'AICc tel que  $AIC = 2k - 2\ln(L)$  avec k, le nombre de paramètres du modèle et L la fonction de vraisemblance du modèle. De la même manière, l'AICc se définit comme  $AICc = AIC + \frac{2k(k+1)}{n-k-1}$  avec n, la taille de l'échantillon.

*Figure n°9 : Tableau comparatif des méthodes selon leur AIC et AICc*

	AIC	AICc
Naïve		
Snaïve		
X13-ARIMA-SEATS	279.8	280.1
STL	397.87	398.37
STS	430.4096	430.7296
Bagged Model	431.0307	444.364

Holt Winter	284.165	284.245
ETS	427.75	441.08
TBATS	455.85	457.7177
ADAM ETS	361.89	377.44
ADAM ETS SARIMA	386.4560	401.9988
SSARIMA	411.33	411.82
CES	340.88	341.73
SARIMA	279.76	280.09

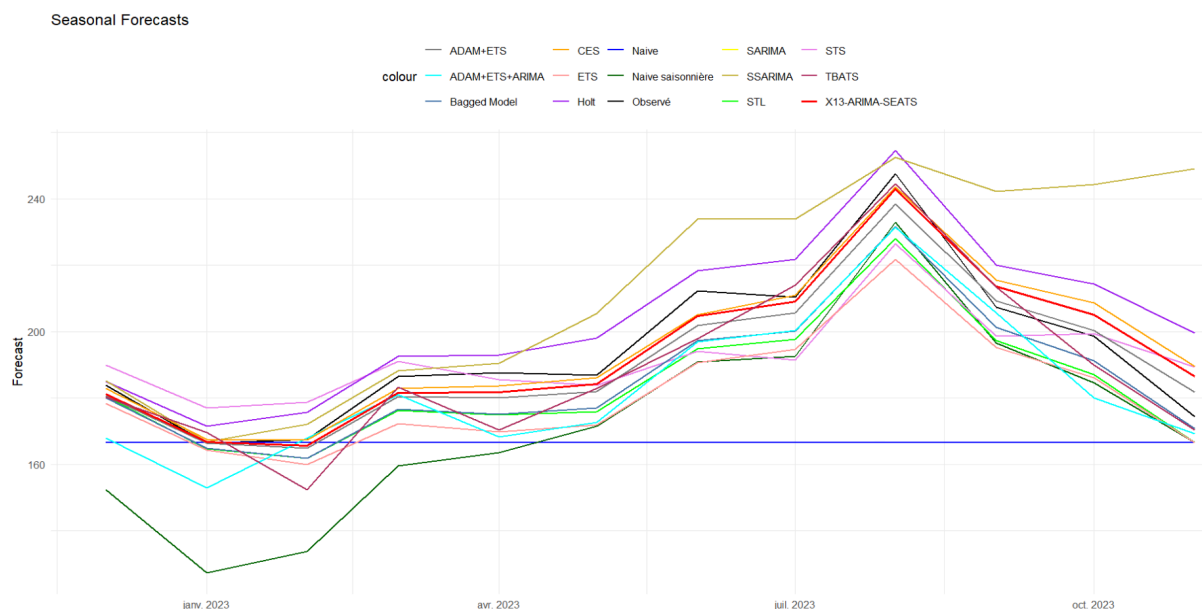
La figure ci-dessous nous permet donc de comparer les différentes méthodes utilisées selon les critères AIC et AICc. On observe ainsi que les critères d'akaike les plus faibles sont pour les méthodes X13-ARIMA-SEATS et SARIMA. En revanche, les méthodes présentant les critères AIC et AICc les plus élevés sont TBATS et STS, indiquant une performance relativement moindre.

## 4.2 Evolution des prévisions

Maintenant que nous avons identifié le meilleur modèle via l'AIC et l'AICc nous devons aussi chercher le modèle qui prédit le mieux nos données. Tout d'abord, le graphique nous permet d'identifier les modèles qui s'éloignent le plus des valeurs observées telles que la méthode naïve qui à une bonne prévision seulement quand la valeur revient vers la dernière valeur observé ce qui est logique, car nous avons une saisonnalité importante. La méthode naïve saisonnière, de son côté une prévision qui suit mieux les variations, mais les premières prévisions sont celles qui sont les plus éloignées des valeurs observées. À partir de mai, elle se rapproche de la valeur réelle mais continue de sous-estimer la prévision en étant l'une des méthodes avec les valeurs les plus basses. Les méthodes ETS, ADAM+ETS ou encore STL suivent les tendances des valeurs observées, mais nous remarquons que les valeurs prédites sont plus faibles que les valeurs observées sur toutes l'années, c'est l'inverse pour les prévisions de holt winters. La méthode SSARIMA fait des bonnes prévisions jusqu'en avril, mais s'éloigne au fur et à mesure pour qu'en septembre les valeurs ne suivent plus la tendance et s'éloigne fortement.

Graphiquement, nous pouvons penser que les meilleures prévisions proviennent des modèles X13-ARIMA-SEATS, CES ou SARIMA.

***Figure n°10 : Graphique des prévisions des chiffres du tourisme en 2023***



## 4.3 Qualité des prévisions

Cependant, pour confirmer l'impression obtenue par le graphique, nous pouvons calculer le MSE et le  $R^2_{OOS}$ , deux mesures d'évaluation des performances des modèles prédictifs. Le MSE est une mesure de l'erreur quadratique moyenne, plus il est faible, meilleure est la performance du modèle, car cela signifie que les prédictions du modèle sont plus proches des valeurs réelles. En revanche, le  $R^2_{OOS}$  est une mesure de la capacité prédictive par rapport à un modèle de référence, dans notre cas la méthode naïve. Une valeur de  $R^2_{OOS}$  inférieure à 0 indique que le modèle est moins performant que la méthode naïve, tandis qu'une valeur élevée indique une bonne capacité à généraliser les prédictions.

***Figure n°11 : Tableau comparatif des méthodes selon leur MSE et  $R^2_{OOS}$***

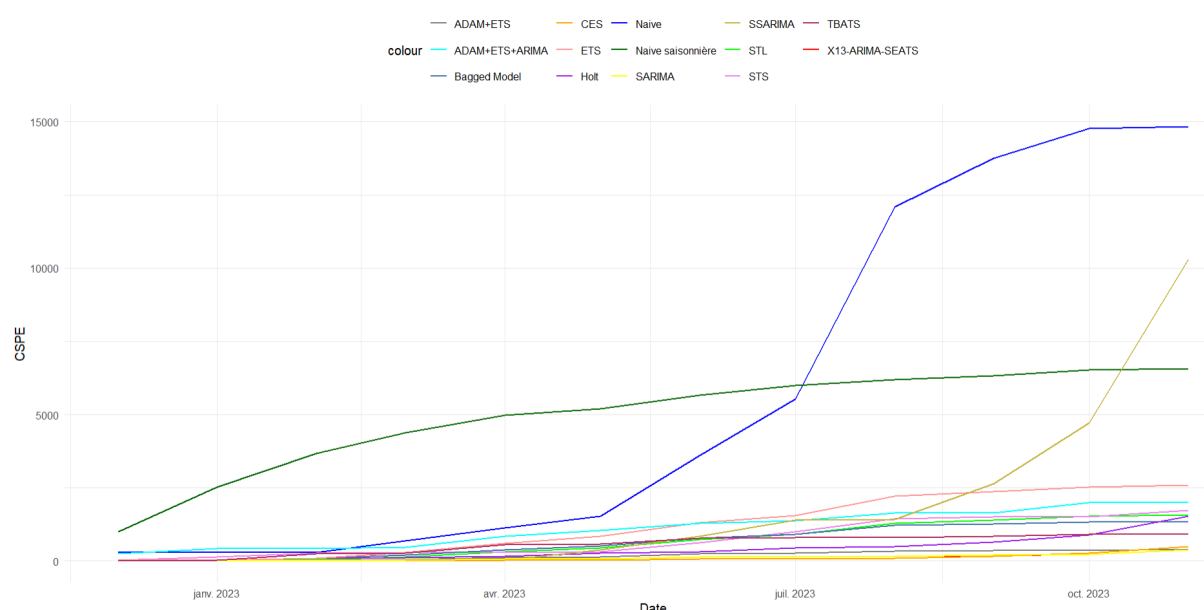
	MSE	$R^2_{OOS}$
SNaïve	1144.393	0
Naïve	332.67	0.7093

X13-ARIMA-SEATS	49.99	0.95631
STL	36.04	0.96850
STS	152.72	0.86655
Bagged Model	42.41182	0.96294
Holt Winter	67.77	0.94078
ETS	44.98	0.96069
TBATS	62.66	0.94524
ADAM ETS	48.13584	0.95794
ADAM ETS SARIMA	0.00511225	0.99999
SSARIMA	150.85	0.86819
CES	38.23	0.96659
SARIMA	37.48	0.967249

D'après ces deux mesures, le modèle ADAM ETS SARIMA se démarque largement, suivi par STL, SARIMA et CES qui affichent des performances très similaires entre eux. Les modèles les moins bons selon ces deux mesures sont la méthode naïve suivie par STS et SSARIMA.

En complément, nous pouvons affiner notre évaluation de la qualité des prévisions en utilisant le graphique CSPE (Cumulative Sum of Prediction Errors), qui offre une représentation visuelle de l'accumulation des erreurs de prédiction au fil du temps. Ce graphique nous permet de mieux comprendre la performance des modèles à travers leur tendance d'erreur cumulative.

**Figure n°12 : Représentation graphique de la somme cumulée des erreurs de prédiction**



Nous constatons ainsi, que le modèle naïf et SSARIMA se distinguent par leurs erreurs importantes en fin de la période prévue. Alors que le modèle naïf saisonnier se distinguent par des erreurs importantes en début de période qui se maintiennent en fin de période. Jusqu'en avril, les erreurs de prédiction restent relativement faibles et tous les modèles affichent une performance similaire sauf pour le modèle naïf saisonnier. Cependant, à partir d'avril, un écart significatif se creuse, mettant en évidence les cinq meilleurs modèles selon cette mesure : X13-ARIMA-SEATS, SARIMA, CES et ADAM ETS SARIMA. Ces observations fournissent des indications supplémentaires sur la fiabilité des prévisions de chaque modèle, renforçant ainsi notre compréhension de leur performance relative.

## 4.4 Précision des prévisions

Maintenant que nous avons identifié les modèles présentant les meilleures estimations et prévisions, nous devons effectuer un dernier test avant de choisir le modèle optimal pour nos données. Ce test final vise à vérifier si les prévisions des modèles diffèrent de manière significative de la méthode naïve saisonnière. L'hypothèse nulle est que le MSE de la méthode naïve saisonnière est égal au MSE

des autres méthodes. Une p-value inférieure à 0,05 indiquerait une différence significative entre les deux MSE, suggérant que l'utilisation d'une méthode plus complexe est plus efficace que la méthode naïve saisonnière.

*Figure n°13 : Tableau comparatif des méthodes selon le test Diebold-Mariano*

	P-value	DM
X13-ARIMA-SEATS	0.003634	3.6787
STL	0.02026	2.7109
STS	0.02216	2.6607
Bagged Model	1	0
Holt Winter	0.02839	2.5217
ETS	0.0712	1.9967
TBATS	1	0
ADAM ETS	1	0
ADAM ETS SARIMA	1	0
SSARIMA	1	0
CES	1	0
SARIMA	1	0

Ainsi, dans notre cas, X13-ARIMA-SEATS, STL, STS et Holt Winter présentent une différence significative par rapport à la méthode naïve saisonnière au seuil de risque de 5 %. La méthode ETS est significativement différente de la méthode naïve saisonnière au seuil de 10 %. Tous les autres modèles ont des valeurs de MSE qui ne diffèrent pas significativement de celle de la méthode naïve saisonnière au seuil de risque de 10 %.

Il est important de noter que nos prévisions ont été établies pour un horizon d'un mois en avance, et non d'un an avec un pas d'un mois. Cette solution a été envisagée, car nous étions en présence d'une variance négative pour nos

méthodes lorsqu'elles étaient utilisées avec la méthode naïve ou la méthode naïve saisonnière. La variance négative indique que la méthode alternative présente une variance inférieure à celle de la méthode de référence (dans ce cas, la méthode naïve saisonnière), suggérant une performance moins stable ou moins précise.

## 4.5 Choix du meilleur modèle

En utilisant différentes mesures telles que l'AIC, l'AICc, le MSE, le  $R^2$ OOS, le CSPE, et le test de Diebold-Mariano (DM), nous sommes en mesure de sélectionner le modèle optimal. Pour ce faire, nous cherchons à maximiser le  $R^2$ OOS, qui reflète la qualité de prédiction du modèle sur des données hors échantillon. En parallèle, nous cherchons à minimiser les autres paramètres, tels que l'AIC, l'AICc, le MSE et le CSPE, car des valeurs plus basses indiquent généralement une meilleure adéquation du modèle aux données et une plus grande précision prédictive. En outre, le test de Diebold-Mariano (DM) nous permet de déterminer si les différences de performance entre les modèles sont statistiquement significatives du modèle naïf, aidant ainsi à confirmer le choix du modèle optimal.

Après avoir examiné les résultats du test DM, nous ne considérons pas ETS comme un modèle optimal, car il n'est pas significatif au seuil de 5 %. Nous constatons que les modèles les plus performants selon le test DM, classés par ordre de performance, sont X13-ARIMA-SEATS, STL, STS et Holt Winter. Les autres modèles présentent un MSE similaire à celui de la méthode naïve saisonnière, ce qui les exclut du choix du meilleur modèle. En analysant les critères AIC et AICc, le classement des modèles les plus favorables est X13-ARIMA-SEATS, Holt Winter, STL et STS. En ce qui concerne le MSE et le  $R^2$ OOS, les meilleurs modèles sont dans l'ordre STL, X13-ARIMA-SEATS, STS et Holt Winter.

Parmi ces quatre modèles potentiels, STS se révèle être le moins performant pour les critères AIC, AICc et le deuxième moins bon selon MSE,  $R^2$ OOS et le test DM, il n'est donc pas le meilleur modèle à utiliser pour notre prévision. En examinant l'évolution des prévisions, on observe que STL présente des prévisions moins précises par rapport à X13-ARIMA-SEATS. En effet, elle sous-estime plus la prévision que X13 qui sous-estime elle aussi la prévision entre janvier et septembre,



bien qu'elle soit plus performante que X13-ARIMA-SEATS selon les critères MSE et le  $R^2_{OOS}$ . La méthode STL n'est donc pas optimale pour la prévision de nos données. De son côté, Holt Winters est un modèle moins performant que X13-ARIMA-SEATS selon tous les critères, choisir ce modèle n'est donc pas pertinent. Au final, le meilleur modèle semble être X13-ARIMA-SEATS, en effet, c'est le seul modèle dont le DM test est significatif au seuil de 5%. De plus, en le comparant aux deux autres modèles, il est le plus performant selon deux critères sur quatre. De plus, graphiquement (figure n°10) on voit clairement que ce modèle suit bien l'évolution des valeurs observées tout en étant relativement proches de valeurs. Enfin, le graphique CSPE (figure n°12), X13-ARIMA-SEATS est un des modèles minimisant le plus l'accumulation des erreurs de prédiction au fil du temps.

## 5 Le meilleur modèle

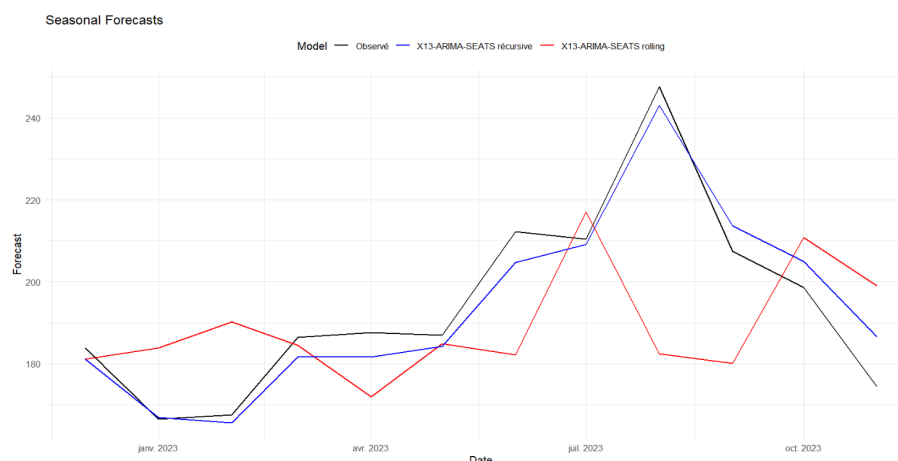
Pour conclure notre analyse, nous allons procéder à la prévision des valeurs sur une année, avec un pas d'un mois, en utilisant notre meilleur modèle X13-ARIMA-SEATS. Nous avons déjà effectué une première prévision récursive sur 12 pas dans la partie précédente. Cela signifie que nous avons prédit les 12 valeurs directement à partir de toutes nos observations. Dans cette partie, nous utiliserons la méthode de la fenêtre roulante à un pas, également appelée rolling à un pas. Cette méthode consiste à prédire les valeurs en utilisant le même nombre d'observations à chaque nouvelle prévision d'une valeur.

*Figure n°14 : Tableau comparatif des méthodes de prévisions avec les valeurs réelles*

	Valeur X13-ARIMA-SEATS (rolling)	Valeur X13-ARIMA-SEAT S (récursive)	Valeur réelle
Décembre 2022	181.1753	181.1753	183.8795
Janvier 2023	183.8221	166.8798	166.4794
Février 2023	190.2117	165.6286	167.5793
Mars 2023	184.4640	181.6431	186.4792
Avril 2023	171.9852	181.6788	187.6792
Mai 2023	184.9073	184.2779	186.9792
Juin 2023	182.1861	204.7080	212.2792
Juillet 2023	217.0703	209.1373	210.4792
Août 2023	182.3772	243.0152	247.5791
Septembre 2023	180.1446	213.6579	207.3791
Octobre 2023	210.7693	205.0098	198.6791
Novembre 2023	199.0469	186.5755	174.3791

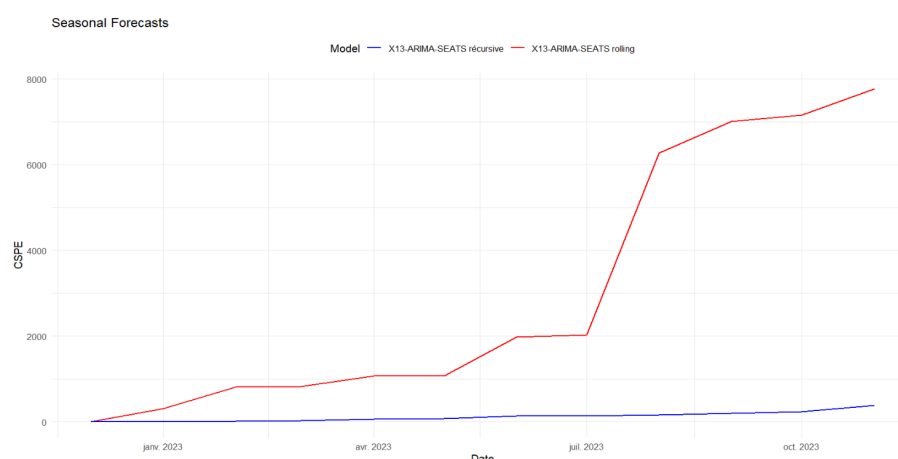
En examinant les résultats obtenus (figure n°14) par les différentes méthodes, nous constatons que les valeurs prédites par la méthode de la fenêtre roulante présentent généralement des écarts plus importants par rapport aux valeurs réelles que celles générées par la méthode réursive.

**Figure n°15 : Evolution des prévisions avec X13-ARIMA-SEATS selon la méthode de prévision**



Cette observation est confirmée visuellement par le graphique (figure n°15), qui met en évidence les différences entre les deux approches et les valeurs observées. En effet, les prévisions de la méthode réursive semblent mieux suivre la saisonnalité des données réelles, ce qui est cohérent avec une accumulation d'erreurs de prédiction (figure n°16) nettement inférieure par rapport à la méthode rolling à 1 pas sur 1 mois.

**Figure n°16 : CSPE de la méthode X13-ARIMA-SEATS selon la méthode de prévision**



En conclusion, l'application de la méthode X13-ARIMA-SEATS avec une prévision réursive à 12 pas produit des résultats remarquablement précis, avec des erreurs minimales. En revanche, la méthode rolling présente des erreurs plus importantes et des prévisions moins fiables. Ces différences peuvent s'expliquer par la disponibilité limitée des données (52 observations), ce qui entraîne une perte d'informations significatives sur la saisonnalité de la série lorsque les premières valeurs sont exclues.

## 6 Conclusion et discussion

Dans cette étude, notre objectif est de faire des prévisions à 1 an du chiffre d'affaires dans le domaine du tourisme. Pour cela, nous avons commencé par procéder à une analyse exploratoire pour connaître les informations principales de notre base de données. Nous avons remarqué que les confinements liés au covid 19 conduisent à deux points atypiques dans notre série, ce qui nous a amenés à corriger notre base avant de procéder à des prédictions à un an. De plus, notre série présente une saisonnalité que nous devons prendre en compte pour l'analyse de nos données.

Nous avons par la suite estimé différents modèles, avec ou sans lissage exponentiel, afin de déterminer celui qui offrait les prévisions les plus fiables pour l'année à venir. A l'aide de diverses mesures telles que l'AIC, l'AICc, le MSE, le  $R^2$ OOS, le CSPE et le test DM, nous avons identifié le modèle optimal. Celui-ci s'est avéré être le modèle fourni par la méthode X13-ARIMA-SEATS, capable de capturer efficacement les fluctuations saisonnières et de produire des prévisions proches de celles attendues.

Cependant, nous avons rencontré des défis méthodologiques qui nécessitent une certaine prudence dans l'interprétation de nos résultats. En particulier, le calcul manuel des indicateurs AIC et/ou AICc (pour les modèles STS, Holt Winter et TBATS) peut introduire une certaine imprécision par rapport aux calculs automatisés effectués par le logiciel R. De plus, la présence de variances négatives lors du test de Diebold et Mariano pour certains modèles a limité la portée de notre analyse, nous obligeant à restreindre le test à un horizon spécifique ( $h=1$ ).

Malgré ces limitations, nos résultats fournissent une base solide pour comprendre les tendances du chiffre d'affaires dans le secteur du tourisme et pour la prévision des données.

## 7 Annexe

### Annexe°1 : Sortie R point atypique

```
> fit
Series: yy
Regression with ARIMA(0,1,0)(0,1,1)[12] errors

Coefficients:
      sma1      LS20      TC21
    -0.7067  -50.9791  -31.1643
s.e.    0.2489    6.8816    6.2375

sigma^2 = 50.96:  log likelihood = -175.04
AIC=358.09  AICc=358.96  BIC=365.82

Outliers:
  type ind   time coefhat  tstat
1  LS  20 2020:03  -50.98 -7.408
2  TC  21 2020:04  -31.16 -4.996
```

### Annexe°2 : Sortie R statistique descriptive

```
> summary(adj)
      y
Min.   : 81.35
1st Qu.:107.83
Median :129.50
Mean    :140.35
3rd Qu.:172.21
Max.    :247.58
```

### Annexe°3 : Sortie R test seasonal dummies

```
> show(sd_ts)
Test used: SeasonalDummies

Test statistic: 48.39
P-value: 0
```

### Annexe°4 : Sortie R test Weibel-Ollech

```
> show(wot)
Test used: WO

Test statistic: 1
P-value: 1.110223e-16 1.887379e-14 4.31955e-07
```

### Annexe°5 : Sortie R test log level

```
> s_transform(myregx13)
tfunction adjust aicdiff
      Auto      None      -2
```

### Annexe°6 : Sortie R décomposition

```
> summary(mysax13$regarima)
y = regression model + arima (0, 1, 0, 0, 1, 1)

Model: RegARIMA - X13
Estimation span: from 8-2018 to 11-2022
Log-transformation: no
Regression model: no mean, no trading days effect, no leap year effect, no Easter effect, no outliers

Coefficients:
ARIMA:
      Estimate Std. Error T-stat Pr(>|t|)
BTheta(1) -0.8695      0.1851 -4.698 3.23e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.071 on 37 degrees of freedom
Log likelihood = -137.9, aic = 279.8, aicc = 280.1, bic(corrected for length) = 4.006
```

### Annexe°7 : Sortie R modèle naïf

```
> summary(prev_naive)

Forecast method: Naive method

Model Information:
Call: naive(y = adj1, h = 12)

Residual sd: 18.2392

Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.4074435 18.23917 15.19092 -0.672803 11.83457 0.6106102 -0.1598431
```

### Annexe°8 : Sortie R modèle STL

```
> summary(prevstl)

Forecast method: STL + ETS(A,N,N)

Model Information:
ETS(A,N,N)

Call:
ets(y = na.interp(x), model = etsmodel, allow.multiplicative.trend = allow.multiplicative.trend)

Smoothing parameters:
  alpha = 0.9999

Initial states:
  l = 101.8479

sigma: 6.1226

      AIC      AICC      BIC
397.8712 398.3712 403.7249

Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 1.576063 6.003665 4.810206 0.9974529 3.904784 0.1933498
      ACF1
Training set 0.05232694
```

### Annexe°9 : Sortie R modèle X13

```
> summary(myregx13)
y = regression model + arima (0, 1, 0, 0, 1, 1)

Model: RegARIMA - X13
Estimation span: from 8-2018 to 11-2022
Log-transformation: no
Regression model: no mean, no trading days effect, no leap year effect, no Easter effect, no outliers

Coefficients:
ARIMA:
      Estimate Std. Error T-stat Pr(>|t|)
BTheta(1) -0.8695      0.1851 -4.698 3.23e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.071 on 37 degrees of freedom
Log likelihood = -137.9, aic = 279.8, aicc = 280.1, bic(corrected for length) = 4.006
```

### Annexe°10 : Sortie R modèle STS

```
> summary(prevsts)

Forecast method: Basic structural model

Model Information:

Call:
StructTS(x = adj1)

Variances:
    level      slope      seas  epsilon 
96.4235    0.4663    4.5644    0.0000 

Error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 1.021157 12.35809 10.02688 0.3428324 8.125495 0.403038
              ACF1
Training set 0.2887748
```

### Annexe°11 : Sortie R modèle bagged model

```
> summary(prevbag)

Forecast method: baggedModel

Model Information:
Series: adj1
Model:  baggedModel
Call:   baggedModel(y = adj1)

Error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 1.837086 6.762608 5.196372 1.010032 3.904976 0.208872
              ACF1
Training set 0.4811567
```

### Annexe°12 : Sortie R modèle SARIMA

```
> summary(prevsarima)

Forecast method: ARIMA(0,1,0)(0,1,1)[12]

Model Information:
Series: adj1
ARIMA(0,1,0)(0,1,1)[12]

Coefficients:
      sma1
      -0.8700
s.e.      1.0107

sigma^2 = 51.29: log likelihood = -137.88
AIC=279.76  AICc=280.09  BIC=283.08

Error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.8615217 6.122107 4.296124 0.4714378 3.376651 0.1726859
              ACF1
Training set 0.03751671
```



### Annexe°13 : Sortie R modèle Holt Winters

```
> summary(prevhw)

Forecast method: HoltWinters

Model Information:
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = adj1)

Smoothing parameters:
alpha: 0.6676997
beta : 0.1103804
gamma: 1

Coefficients:
      [,1]
a 184.815772
b   2.744801
s1 -2.519941
s2 -18.824909
s3 -17.421730
s4 -3.141564
s5 -5.675016
s6 -3.197160
s7 14.328800
s8 15.030268
s9 44.946023
s10 7.774247
s11 -0.634795
s12 -18.136154

Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.942526 8.232479 6.281808 0.4982585 4.899191 0.2525019 0.167619
```

### Annexe°14 : Sortie R modèle ETS

```
> summary(prevets)

Forecast method: ETS(M,N,A)

Model Information:
ETS(M,N,A)

Call:
ets(y = adj1)

Smoothing parameters:
alpha = 0.9999
gamma = 1e-04

Initial states:
l = 120.8605
s = 13.7678 9.7098 -9.0934 -11.1091 -8.7167 -20.9192
    -16.7002 -2.7173 -14.3039 5.1295 14.2349 40.7179

sigma: 0.0604

      AIC      AICC      BIC
427.7493 441.0826 457.0180

Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 1.156319 6.706359 5.379111 0.6929986 4.215209 0.2162174 -0.01607618
```

### Annexe°15 : Sortie R modèle TBATS

```
> summary(prev_tbats)

Forecast method: TBATS(0.262, {0,0}, 0.854, {<12,5>})

Model Information:
TBATS(0.262, {0,0}, 0.854, {<12,5>})

Call: tbats(y = adj1)

Parameters
  Lambda: 0.262138
  Alpha: 0.3431267
  Beta: 0.1940131
  Damping Parameter: 0.853544
  Gamma-1 Values: -0.002769191
  Gamma-2 Values: 0.01505411

Seed States:
      [,1]
[1,] 9.072470257
[2,] 0.083813214
[3,] 0.549207863
[4,] 0.213990792
[5,] 0.183850256
[6,] -0.058622440
[7,] 0.147767792
[8,] -0.080564246
[9,] -0.021771227
[10,] -0.001894535
[11,] 0.103709053
[12,] -0.089231871
attr(,"lambda")
[1] 0.2621383

Sigma: 0.2242408
AIC: 455.851

Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.8664827 7.915939 6.068458 0.4208068 4.909208 0.2439262
      ACF1
Training set -0.002857725
```

### Annexe°16 : Sortie R modèle ADAM ETS

```
> fitadam1 <- auto.adam(adj1, model="ZZZ", lags=c(1,12), select=TRUE)
> fitadam1
Time elapsed: 19.23 seconds
Model estimated using auto.adam() function: ETS(ANA)
Distribution assumed in the model: Normal
Loss function type: likelihood; Loss function value: 167.0658
Persistence vector g:
alpha gamma
  1      0

Sample size: 52
Number of estimated parameters: 15
Number of degrees of freedom: 37
Information criteria:
      AIC      AICc      BIC      BICc
364.1315 377.4649 393.4002 419.7418
```

### Annexe°17 : Sortie R modèle ADAM ETS SARIMA

```
> fitadam2 <- auto.adam(adj1, model="ZZN", lags=c(1,12), orders=list(ar=c(3,3), i=(2), ma=c(3,3), select=TRUE))
> fitadam2
Time elapsed: 29.48 seconds
Model estimated using auto.adam() function: ETS(MAN)+SARIMA(1,0,0)[12]
Distribution assumed in the model: Generalised Normal with shape=0.273
Loss function type: likelihood; Loss function value: 166.5354
Persistence vector g:
  alpha  beta
0.9736 0.0065

ARMA parameters of the model:
AR:
phil[12]
  0.8268

Sample size: 52
Number of estimated parameters: 19
Number of degrees of freedom: 33
Information criteria:
      AIC      AICc      BIC      BICc
371.0708 394.8208 408.1445 455.0655
```

### Annexe°18 : Sortie R modèle SSARIMA

```
> summary(fitssarima)
Time elapsed: 3.18 seconds
Model estimated: SARIMA(0,1,0)[1](0,0,2)[12]
Matrix of MA terms:
      Lag 12
MA(1) 0.6677
MA(2) 0.7524
Initial values were produced using backcasting.

Loss function type: likelihood; Loss function value: 202.6625
Error standard deviation: 12.2819
Sample size: 52
Number of estimated parameters: 3
Number of degrees of freedom: 49
Information criteria:
      AIC      AICc      BIC      BICc
411.3250 411.8250 417.1787 418.1665
```

### Annexe°19 : Sortie R modèle CES

```
> summary(fitces)
Time elapsed: 1.01 seconds
Model estimated: CES(p)
a0 + ia1: 1.9416+1.0192i
b: -0.0054
Initial values were produced using backcasting.

Loss function type: likelihood; Loss function value: 166.4377
Error standard deviation: 6.183
Sample size: 52
Number of estimated parameters: 4
Number of degrees of freedom: 48
Information criteria:
      AIC      AICc      BIC      BICc
340.8755 341.7265 348.6804 350.3618
```

## 8 Bibliographie

Atout france, “Note de conjoncture - Mars 2024” :

<https://www.atout-france.fr/fr/actualites/note-de-conjoncture-mars-2024>

Campus France, “Tourisme ; 2023 une année record pour la France” :

<https://www.campusfrance.org/fr/actu/tourisme-2023-une-annee-record-pour-la-france#:~:text=Car%2C%20rappelle%20le%20site%20pr%C3%A9sidentiel,%2C5%25%20du%20PIB%20fran%C3%A7ais>

Eurostat, “Chiffre d'affaires dans les services - données mensuelles” :

[https://ec.europa.eu/eurostat/databrowser/view/sts\\_setu\\_m\\$dv\\_667/default/table?lang=en&category=tour.tour\\_inda.tour\\_sts](https://ec.europa.eu/eurostat/databrowser/view/sts_setu_m$dv_667/default/table?lang=en&category=tour.tour_inda.tour_sts)

Le portail de la direction général des entreprises :

<https://www.entreprises.gouv.fr/fr/tourisme>

# 9 Table des matières

1 Introduction.....	2
2 Analyse exploratoire.....	3
2.1 Présentation de la série.....	3
2.2 Détection des points atypiques.....	4
2.3 Statistiques descriptives.....	5
2.4 Détection de la saisonnalité.....	7
2.5 Désaisonnalisation et décomposition.....	9
3 Estimation et prévision.....	11
3.1 Méthode de prévision sans lissage exponentiel.....	11
3.1.1 Méthode naïve et naïve saisonnière.....	11
3.1.2 Méthode non paramétrique : STL.....	11
3.1.3 Méthode semi-paramétrique : X13-ARIMA- SEATS.....	12
3.1.4 Méthode paramétrique : STS.....	12
3.1.5 Bagged Model.....	13
3.1.6 SARIMA (p,q,d)(P,D,Q) <sub>12</sub> .....	13
3.2 Méthode de prévision avec lissage exponentiel.....	14
3.2.1 Holt-Winters.....	14
3.2.2 ETS.....	15
3.2.3 TBATS.....	15
3.2.4 ADAM ETS.....	15
3.2.5 ADAM ETS SARIMA.....	16
3.2.6 SSARIMA.....	16
3.2.7 CES.....	16
4 Comparaisons des modèles.....	17
4.1 Meilleur modèle estimé.....	17
4.2 Evolution des prévisions.....	18
4.3 Qualité des prévisions.....	19
4.4 Précision des prévisions.....	21
5 Le meilleur modèle.....	25
6 Conclusion et discussion.....	28
7 Annexe.....	29
8 Bibliographie.....	35
9 Table des matières.....	36