

Projet :

---

## Application de la Méthode de Box-Jenkins Sur une Série Temporelle

---



**PRESENTE PAR :**  
CISSE Pape Waly  
KHLAIF mohamed

**ENCADRE PAR :**  
PR.FADOUA BADAQUI

**ANNEE SCOLAIRE : 2023-2024**

# Table des matières

<b>1</b>	<b>Description</b>	<b>2</b>
<b>2</b>	<b>Analyse des données</b>	<b>2</b>
<b>3</b>	<b>Methode de Box-Jenkins</b>	<b>4</b>
<b>4</b>	<b>Evaluation du modèle</b>	<b>8</b>

# 1 Description

Notre étude sur l'indice de production harmonisé de production industrielle du Sénégal .cet indice permet de mesurer le niveau de production industrielle d'un pays . il peut être utilisé pour analyser les tendances économiques, évaluer la croissance industrielle, suivre les cycles économiques et fournir des indications sur la santé générale de l'économie d'un pays.Nous les avons récupéré du site de L'ANSD(Agence National de la Statistique et de la Demographie).

## 2 Analyse des données

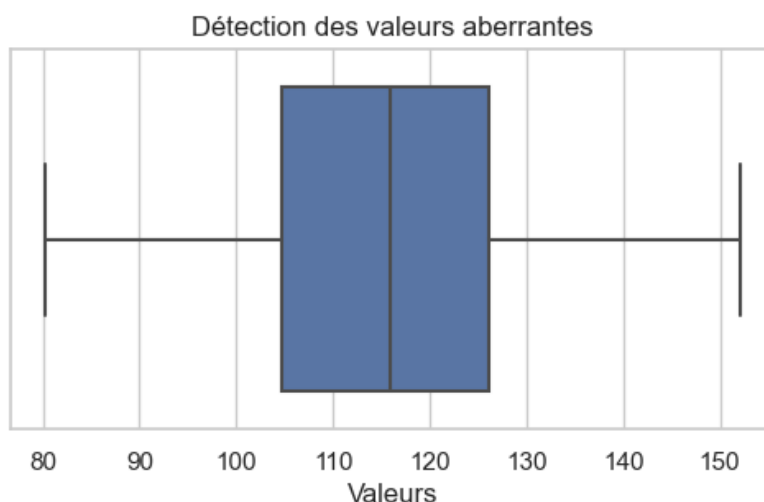
Dans cette il s'agira d'analyser les données afin de voir quel modèle choisir(multiplicatif ou additif),de tester la stationnarité de notre série ,la stationnariser dans le cas il ne l'est pas.Mais avant tous cela on va voir si nos données ne possèdent pas de valeurs manquantes ou aberrantes.

### Valeurs manquantes

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 104 entries, 2015-01-01 to 2023-08-01
Data columns (total 1 columns):
#   Column  Non-Null Count  Dtype  
---  -
0    Value   104 non-null      float64
dtypes: float64(1)
memory usage: 1.6 KB
```

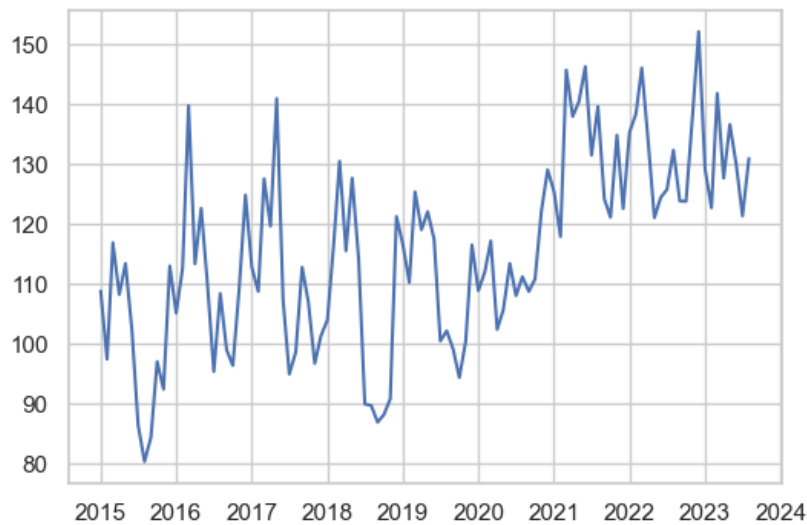
On peut ainsi voir que le nombre de données d'entrées est équivalent aux nombres de données non nulle .de ce fait pas de données manquantes.

### Valeurs aberrantes



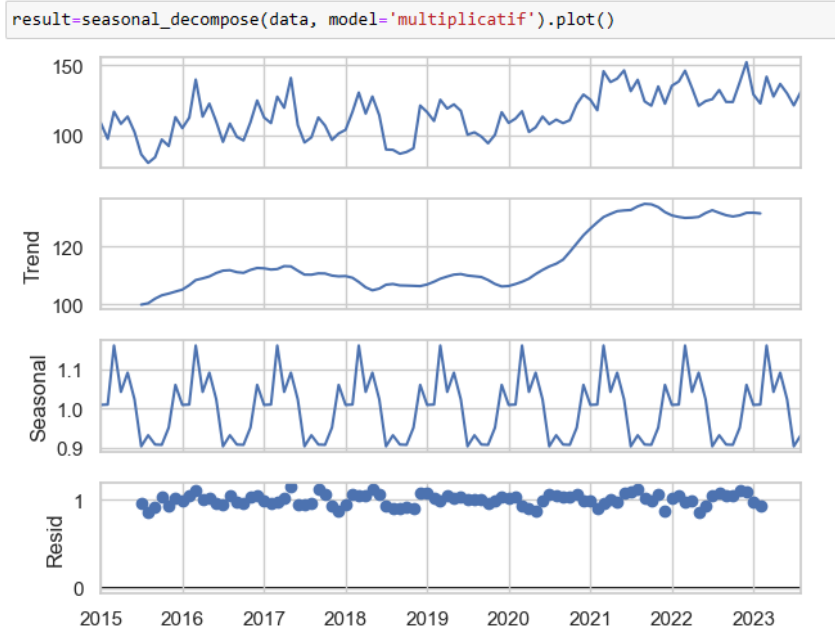
le boxplot ainsi tracé nous permet de confirmer que nos données ne possèdent pas de valeurs aberrantes.

## figures de notre série



On peut voir une saisonnalité dans notre série ainsi tracé ce qu'on va confirmer avec la décomposition de la série.

## Decompistion de la Série :



Vu le diagramme des résidus on peut considérer cette série de modèle multiplicatif on voit bien que les résidus sont de d'espérance et variance constante(homoscedasticité).

## Test de Dickey-fuller augmenté :Stationnarité

Statistique ADF : -1.1027107585242857  
p-value : 0.7139896761599561

Nous pouvons constater que le pvalue est supérieur au seuil de 0.05 donc notre série n'est pas stationnaire. Ainsi on a effectué une différenciation non saisonnière pour la stationnariser.

**La série après différenciation : Stationnarité**



**Test de Dickey-fuller augmenté : Stationnarité**

Statistique ADF : -3.9766925616124325  
p-value : 0.0015375954238151836

Ainsi notre pvalue est inférieur à 0.05 donc notre série est stationnaire.

### 3 Méthode de Box-Jenkins

La méthodologie de Box-Jenkins est une approche couramment utilisée pour modéliser et prévoir les séries temporelles. Elle comprend les étapes suivantes :

#### 1. Identification du modèle

Identification des ordres : Utiliser les graphiques ACF (fonction d'autocorrélation) et PACF (fonction d'autocorrélation partielle) pour déterminer les ordres  $p$ ,  $d$  et  $q$  du modèle ARIMA.

Estimation des paramètres : Utiliser les méthodes d'estimation (telles que la méthode des moindres carrés) pour estimer les paramètres du modèle ARIMA.

#### 2. Validation du modèle

Vérifier si les résidus du modèle ARIMA sont un bruit blanc (c'est-à-dire s'ils ne présentent pas de structure ou de corrélation significative).

Réajustement : Si le modèle ne satisfait pas les critères de bruit blanc, ajuster les ordres du modèle ARIMA et répéter les étapes précédentes.

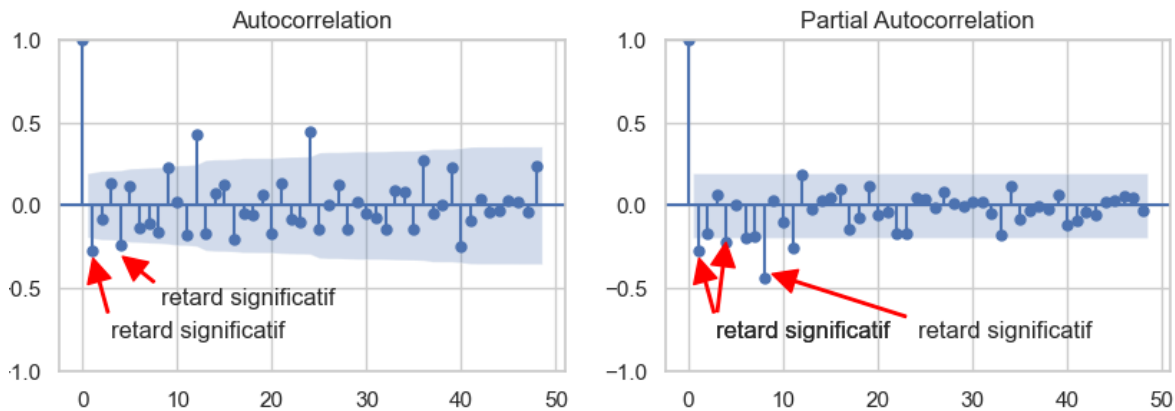
Validation : Valider les performances du modèle en effectuant des prédictions sur des données de validation ou en utilisant des mesures d'évaluation telles que l'erreur quadratique moyenne (RMSE) ou le critère d'information d'Akaike (AIC).

### 3. Application à nos données

#### choix du modèle

Pour identifier le modèle, on doit utiliser les graphes d'autocorrélation (pour les MA) et d'autocorrélation partielle (pour les AR).

Voici nos diagrammes :



On peut voir les retards significatifs au niveau de chaque diagramme. Au vu de nos résultats, on peut prendre comme  $p_{\max}=3$  et  $q_{\max}=2$ . Ainsi, on va choisir le meilleur modèle entre :  $\text{ARIMA}(p,1,q)$  étant donné qu'on a fait une seule différenciation non saisonnière avec  $p$  (0,..., $p_{\max}$ ) et  $q$  (0,..., $q_{\max}$ ).

Sous Python, on a la bibliothèque `pmdarima` qui permet de faire le test ces modèles et de choisir le meilleur d'entre eux sous le critère AIC ou BIC, etc. Dans notre cas, on a utilisé AIC.

```

ARIMA(2,1,1)(2,0,0)[12]      : AIC=667.908, Time=0.27 sec
ARIMA(1,1,1)(2,0,0)[12]      : AIC=659.173, Time=0.27 sec
ARIMA(1,1,1)(1,0,0)[12]      : AIC=666.347, Time=0.13 sec
ARIMA(1,1,1)(2,0,1)[12]      : AIC=661.151, Time=0.34 sec
ARIMA(1,1,1)(1,0,1)[12]      : AIC=660.496, Time=0.22 sec
ARIMA(0,1,1)(2,0,0)[12]      : AIC=664.237, Time=0.10 sec
ARIMA(1,1,2)(2,0,0)[12]      : AIC=661.094, Time=0.47 sec
ARIMA(0,1,0)(2,0,0)[12]      : AIC=668.427, Time=0.09 sec
ARIMA(0,1,2)(2,0,0)[12]      : AIC=660.812, Time=0.19 sec
ARIMA(2,1,2)(2,0,0)[12]      : AIC=661.189, Time=0.42 sec

```

```

Best model:  ARIMA(1,1,1)(2,0,0)[12]
Total fit time: 7.650 seconds

```

ARIMA
ARIMA(1,1,1)(2,0,0)[12]

Maintenant qu'on choisi le modèle on passe à la Validation.

### Validation du modèle

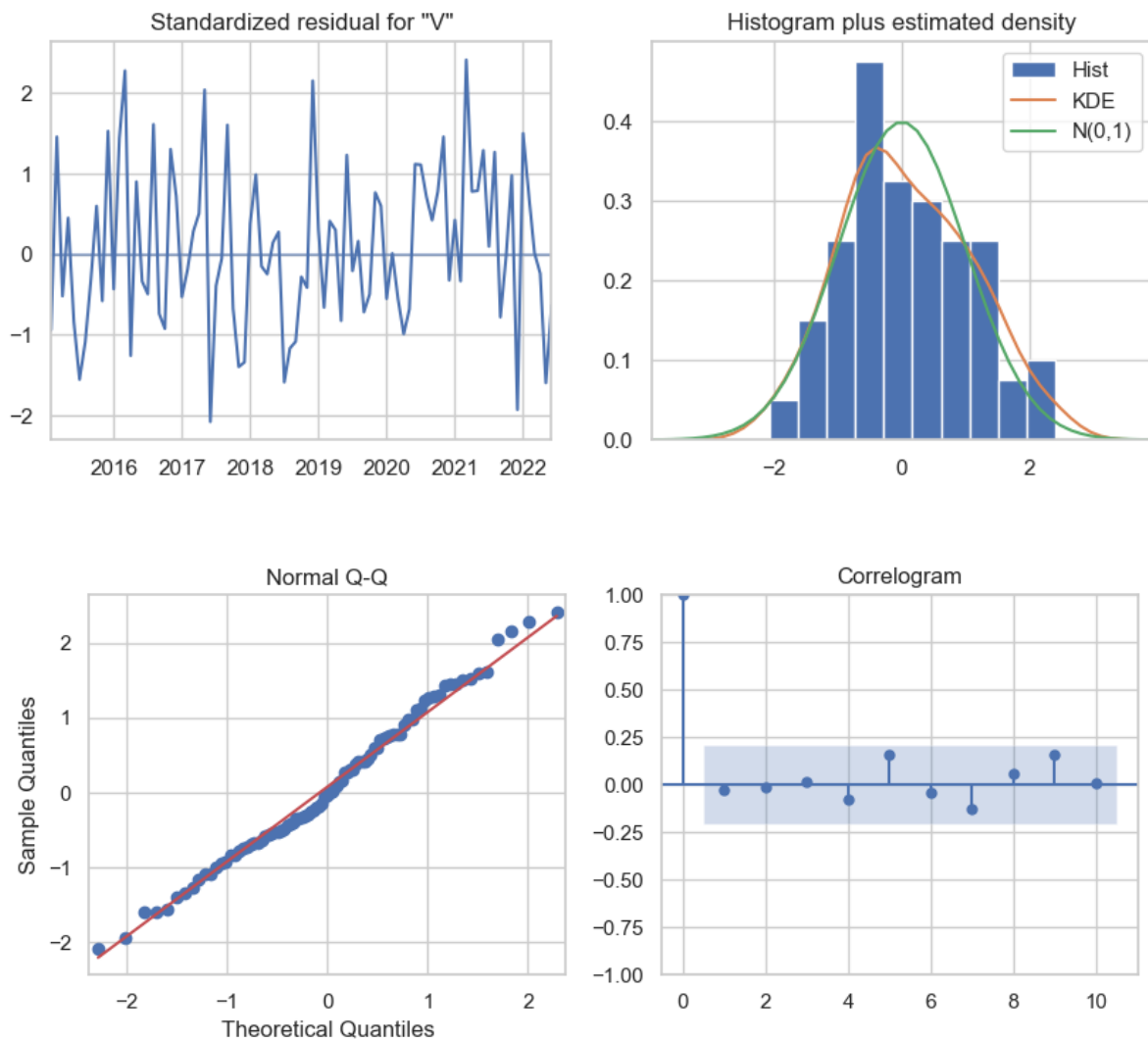
Dans cette partie on va voir si nos résidus sont des bruits blancs et si tous nos paramètres estimés sont significatifs.

```

=====
SARIMAX Results
=====
Dep. Variable:          Value    No. Observations:          98
Model:          SARIMAX(1, 1, 1)x(2, 0, [], 12)    Log Likelihood          -353.994
Date:          Tue, 19 Dec 2023    AIC          717.989
Time:          16:47:18    BIC          730.862
Sample:          01-01-2015    HQIC          723.194
              - 02-01-2023
Covariance Type:          opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ar.L1          0.5239      0.126      4.148      0.000      0.276      0.771
ma.L1         -0.8906      0.067     -13.275      0.000     -1.022     -0.759
ar.S.L12        0.2610      0.105      2.483      0.013      0.055      0.467
ar.S.L24        0.4382      0.115      3.812      0.000      0.213      0.663
sigma2         79.3226     14.452      5.489      0.000     50.996     107.649
=====
Ljung-Box (L1) (Q):          0.00    Jarque-Bera (JB):          1.18
Prob(Q):          0.98    Prob(JB):          0.55
Heteroskedasticity (H):      0.93    Skew:          0.13
Prob(H) (two-sided):      0.83    Kurtosis:          2.52
=====

```

Ainsi on peut voir que tous nos paramètres estimés sont significatifs vu qu'on a un pvalue<0.05 .Aussi le test de Ljung-Box a un pvalue de 0.98 ce qui permet de valider l'hypothèse nulle qui suppose que nos résidus sont nos corrélés.toutefois on va faire un diagnostique plus étendu sur ces résidus pour valider le modèle .



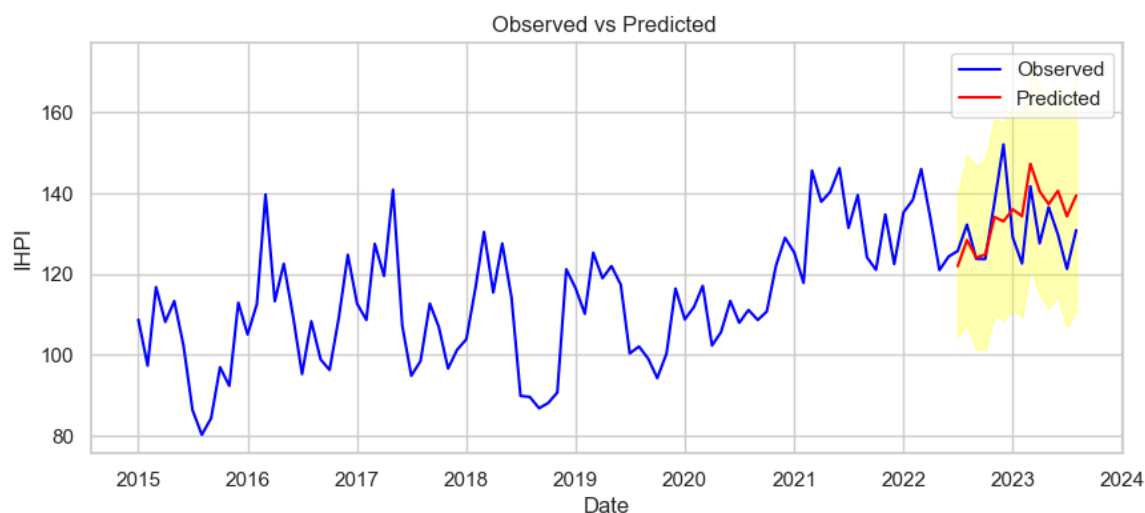
le Q-Q plot et l'histogramme nous permettent de confirmer que nos résidus suivent une loi normale .le corrélogramme des résidus nous confirment que nos résidus sont non corrélés(à part le premier retard tous les autres sont à l'intérieur de l'intervalle de confiance ce qui signifie qu'ils ne sont pas significatifs ).

### Cross-validation :Prédiction

On avait divisé nos données en trainset and testset ;ensuite on a entrainé le modèle sur nos données d'entrainement et finalement valider avec le prédiction sur les données de test.

voici le résultat :





nos prédictions suivent à peu près nos données réelles .ce fait ajouté aux deux autres étapes nous permet de valider notre modèle.

## 4 Evaluation du modèle

pour l'évaluation du modèle on va se baser sur le MAD,MSE,et le MAPE.

	Value	predicted_mean	erreur	erreurabsolu	squarederror	ratioErreur
2022-07-01	125.716038	121.972810	-3.743227	3.743227	14.011751	0.029775
2022-08-01	132.268702	128.427416	-3.841286	3.841286	14.755476	0.029042
2022-09-01	123.784622	124.104477	0.319855	0.319855	0.102307	0.002584
2022-10-01	123.730712	124.848159	1.117448	1.117448	1.248689	0.009031
2022-11-01	137.826682	134.134696	-3.691986	3.691986	13.630764	0.026787
2022-12-01	152.092380	133.075027	-19.017353	19.017353	361.659697	0.125038
2023-01-01	129.108219	136.036581	6.928363	6.928363	48.002211	0.053663
2023-02-01	122.652648	134.352403	11.699755	11.699755	136.884261	0.095389
2023-03-01	141.750052	147.320239	5.570187	5.570187	31.026980	0.039296
2023-04-01	127.621628	140.496885	12.875257	12.875257	165.772238	0.100886
2023-05-01	136.571718	137.315362	0.743644	0.743644	0.553007	0.005445
2023-06-01	129.804821	140.632618	10.827797	10.827797	117.241183	0.083416
2023-07-01	121.297085	134.324020	13.026935	13.026935	169.701041	0.107397
2023-08-01	130.889249	139.453365	8.564116	8.564116	73.344085	0.065430

	Métrique	Ensemble de test
0	MAD	7.283372
1	MSE	81.995264
2	Mape	0.055227

la valeur moyenne de nos erreurs prédictions(mape) est de 0.055 ce qui peut être considérée comme faible vue le nombre d'éléments qu'on a prédit. de même pour l'erreur absolue de deviation. Ainsi le modèle peut être considéré comme performant.