

PROJET PYTHON

Une analyse de l'évolution de l'inflation en France entre
1990 et 2024

Réalisé par :

- MAHI ELAMINE GUENDOUZ - 12306918
- JALAKSHANA KANNAN - 12309965

Table des matières :

Introduction	02
Partie 01 : Importation, Nettoyage, préparation et visualisation des données	02
Phase d'importation	02
Phase de nettoyage et de préparation	03
Phase de visualisation	05
Partie 02 : Analyse de donnée.....	08
Parie 03 : Application du modèle ARIMA.....	12
Analyse de la série temporelle	13
Prédiction pour 2024	14
Partie 04 : Comparaison des indices de prix à la consommation harmonisé de plusieurs pays entre janvier 2019 et décembre 2023	15
Conclusion	16
Annexe	17
Sitographie	19
Table des illustrations	20

Introduction :

Au cours de ses dernières années, tout le monde a remarqué une augmentation vertigineuse des prix des biens et des services, une perte du pouvoir d'achat des ménages et une dépréciation de l'euro face au dollars américain, partout en Europe et plus précisément en France. Ce phénomène due à plusieurs changements dans la sphère politique et économique mondiale, est appelé l'inflation. C'est pour cela, on a voulu étudier et analyser l'inflation en France entre 1990 et 2024 et essayer de trouver des explications à ces changements.

Dans l'optique de réaliser ce travail, on s'est basé sur 6 data bases (trouvées sur le site de l'INSEE) :

- df_1 : Indice_prix_consommation_90_24.csv : Ce fichier contient les indices de prix à la consommation hors tabac pour la période qui s'étend du 01/01/1990 au 01/12/2023.
- df_2 : Indice_prix_consommation_90_24_alimentation.csv : Ce fichier contient les indices de prix à la consommation – produit alimentaire entre 01/01/1990 et 01/12/2023.
- df_3 : Indice_prix_consommation_90_24_energie.csv : Ce fichier contient les indices de prix à la consommation – énergie entre 01/01/1990 et 01/12/2023.
- df_4 : Indice_prix_consommation_90_24_produit_manufacturé.csv : Ce fichier contient les indices de prix à la consommation – produits manufacturés du 01/01/1990 au 01/12/2023.
- df_5 : Taux d'inflation annuel.xlsx : Ce fichier contient les taux d'inflation annuel en France entre 1991 et 2023.
- df_6 : prc_hicp_midx_page_tabular.tsv : Ce fichier contient les indices de prix à la consommation harmonisé de plusieurs pays entre janvier 2019 et décembre 2023.

Partie 01 : Importation, Nettoyage, préparation et visualisation des données :

1- Phase d'importation :

D'abord, on a commencé par importer les bibliothèques nécessaires pour mener à bien cette étude, les bibliothèques sont :

Tableau 1 : Liste des bibliothèques utilisées

import pandas as pd	Pandas permet l'analyse et la manipulation des données en python.
import numpy as np	Numpy permet de réaliser des calculs numériques en python.
import matplotlib.pyplot as plt	Matplotlib.pyplot permet la visualisation des données en python.
import os	Os permet les interactions avec le système d'exploitation.
import statsmodels.api as sm	Statsmodels.api permet de réaliser des estimations statistiques en python.
import seaborn as sns	Seaborn permet la visualisation des données.
from statsmodels.tsa.arima.model import ARIMA	ARIMA permet l'estimation du modèle autoregressive integrated moving average.

```
from sklearn.metrics import  
mean_squared_error
```

Sklearn permet de réaliser l'apprentissage automatique.

Ensuite, on a déclaré le chemin à notre répertoire dans lequel se trouve toutes les data bases avec lesquelles on va travailler. Notre répertoire :

```
Directory_path = r"C:\Users\DELL\OneDrive\Bureau\py2"
```

Enfin, on a importé nos bases de données (csv et xlsx), en utilisant les fonctions « read_csv » et « read_excel » de la bibliothèque pandas et en déclarent le répertoire grâce à la fonction « path.join » du module os. (Comme on peut voir sur la figure 14 en annexe).

2- Phase de nettoyage et de préparation :

Dans cette étape, on a procédé comme suit :

- On a commencé par sélectionner les données dont lesquelles on aurait besoin pour la réalisation de l'analyse et on a supprimer le reste.
- On a aussi renommé les colonnes qu'on a gardé. (Voir figure 15 en annexe)
- Ensuite, on a fusionné les 4 premières data frame, qui contiennent les indices mensuels de prix à la consommation générale hors tabac, alimentaire, de l'énergie et des produits manufacturés dans une seule data frame appelée « IPC_df » (voir figure 16 en annexe) et on a obtenu le tableau suivant :

Tableau 2 : IPC_df DATA

	Date	indice hors tabac	indice alimentation	indice énergie	indice produit manufacturés
0	1990-01	68.09	67.57	52.06	91.96
1	1990-02	68.23	67.72	51.83	92.13
2	1990-03	68.4	67.72	51.69	92.54
3	1990-04	68.78	68.57	51.55	92.74
4	1990-05	68.89	68.57	51.49	92.91
...
403	2023-08	118.0	131.29	157.99	104.54
404	2023-09	117.37	130.92	161.12	105.26
405	2023-10	117.54	130.97	160.3	105.43
406	2023-11	117.33	131.47	158.53	105.47
407	2023-12	117.5	131.52	156.66	105.17

408 rows x 5 columns

- Par la suite, on a opéré une vérification de la composition de « IPC_df », pour vérifier que la date est sous format Datetime et les indices sous format numérique (flotteur) (voir figure 04 en annexe), ce qui n'était pas le cas, donc on a exécuté une conversion des variables. (Voir figure 17 en annexe).
- Puis, à partir de la data frame « IPC_df », on a créé une nouvelle data frame nommée « IPC_df_annuelle », qui contient les taux d'inflation annuel hors tabac, alimentaire, énergétique et des produits manufacturés de la France entre 1991 et 2023. (Voir figure 18 en annexe). On a calculé le taux d'inflation annuel tel que :

Équation 1 : Formule de calcul du taux d'inflation annuel

$$\text{Inflation rate}(t) = \frac{\text{CPI}(t) - \text{CPI}(t-1)}{\text{CPI}(t-1)} \times 100$$

CPI_t : Indice de prix à la consommation de l'année t .

CPI_{t-1} : Indice de prix à la consommation de l'année précédente $t - 1$.

On a eu comme résultat le tableau suivant :

Tableau 3 : IPC_df_annuelle DATA

	Year	Taux d'inflation hors tabac	Taux d'inflation alimentation	Taux d'inflation énergie	Taux d'inflation produits manufacturés
0	1991	2.429322	1.783384	-0.961887	2.427905
1	1992	1.620050	-0.707314	-0.612131	1.492073
2	1993	1.534909	-0.437853	3.692652	1.247061
3	1994	1.284625	1.036343	-0.179791	0.770635
4	1995	1.739360	0.389213	2.295782	1.780475
5	1996	1.488480	0.951593	5.291187	1.101211
6	1997	0.916964	0.901022	-1.036102	1.190714
7	1998	0.544649	0.773024	-4.846394	1.123145
8	1999	1.568032	0.420665	10.615304	1.036113
9	2000	1.682243	2.598929	6.558654	1.011517
10	2001	1.786154	4.198811	-2.826421	1.671170
11	2002	1.833756	-0.345856	3.857708	1.249504
12	2003	1.587302	1.482624	-2.190760	1.858662
13	2004	2.090877	-0.499593	9.045681	1.688015
14	2005	2.197676	0.677254	10.180231	0.882061
15	2006	1.611631	1.293452	-0.025126	1.013917
16	2007	2.882582	2.739259	11.400734	1.766291
17	2008	1.126394	1.777923	-7.763380	1.602309
18	2009	1.247580	-0.884107	3.359141	1.412654
19	2010	1.937819	0.223547	9.717314	1.749475
20	2011	2.648937	3.446442	5.416321	2.574336
21	2012	1.586814	2.208753	0.211173	2.142007
22	2013	1.175160	0.452580	-0.705882	1.779501
23	2014	0.604473	-0.340681	-5.160259	1.832092
24	2015	1.203844	0.482170	-2.692035	2.951088
25	2016	1.604926	0.820574	5.556725	2.064595
26	2017	1.344488	0.889328	0.505725	2.297200
27	2018	1.465526	2.216148	3.183385	1.732978
28	2019	1.675270	1.558491	3.925234	1.884073
29	2020	0.144314	0.624301	-7.422241	1.249871
30	2021	2.676516	0.977860	16.517332	1.639344
31	2022	5.703635	11.456346	11.938383	5.640399
32	2023	3.196908	5.426854	1.760312	2.444964

- Ensuite, on a créé aussi une data frame appelée « IPC_df_mensuelle » à partir de « IPC_df », qui contient les taux d'inflation mensuel hors tabac, alimentaire, énergétique et des produits manufacturés de la France entre 1990 et 2023. (Voir figure 19 en annexe)

On a calculé le taux d'inflation mensuel de cette manière :

Équation 2 : Formule de calcul du taux d'inflation mensuel

$$\text{Inflation rate}(t) = \frac{\text{CPI}(t) - \text{CPI}(t-1)}{\text{CPI}(t-1)} \times 100$$

CPI_t : Indice de prix à la consommation du mois t.

CPI_{t-1} : Indice de prix à la consommation du mois précédent t – 1.

Tableau 4 : IPC_df_mensuelle DATA

	Date	taux inflation mensuel hors tabac	taux inflation mensuel alimentation	taux inflation mensuel énergie	taux inflation mensuel produit manufacturés
1	1990-02-01	0.205610	0.221992	-0.441798	0.184863
2	1990-03-01	0.249157	0.000000	-0.270114	0.445023
3	1990-04-01	0.555556	1.255168	-0.270845	0.216123
4	1990-05-01	0.159930	0.000000	-0.116392	0.183308
5	1990-06-01	-0.014516	-0.700015	-0.466110	0.150683
...
403	2023-08-01	1.018748	0.343932	6.605938	1.485293
404	2023-09-01	-0.533898	-0.281819	1.981138	0.688732
405	2023-10-01	0.144841	0.038191	-0.508937	0.161505
406	2023-11-01	-0.178663	0.381767	-1.104180	0.037940
407	2023-12-01	0.144890	0.038031	-1.179587	-0.284441

3- Phase de visualisation :

En utilisant, la bibliothèque matplotlib et plus précisément le sous-module pyplot, on a représenté les courbes de l'évolution des taux d'inflation annuels et mensuels générale hors tabac, alimentaire, énergétique et des produits manufacturés entre 1991 et 2024 ; et l'évolution du taux d'inflation annuel général de la France (data frame : df_5) entre 1991 et 2023. (Voir figure en annexe).

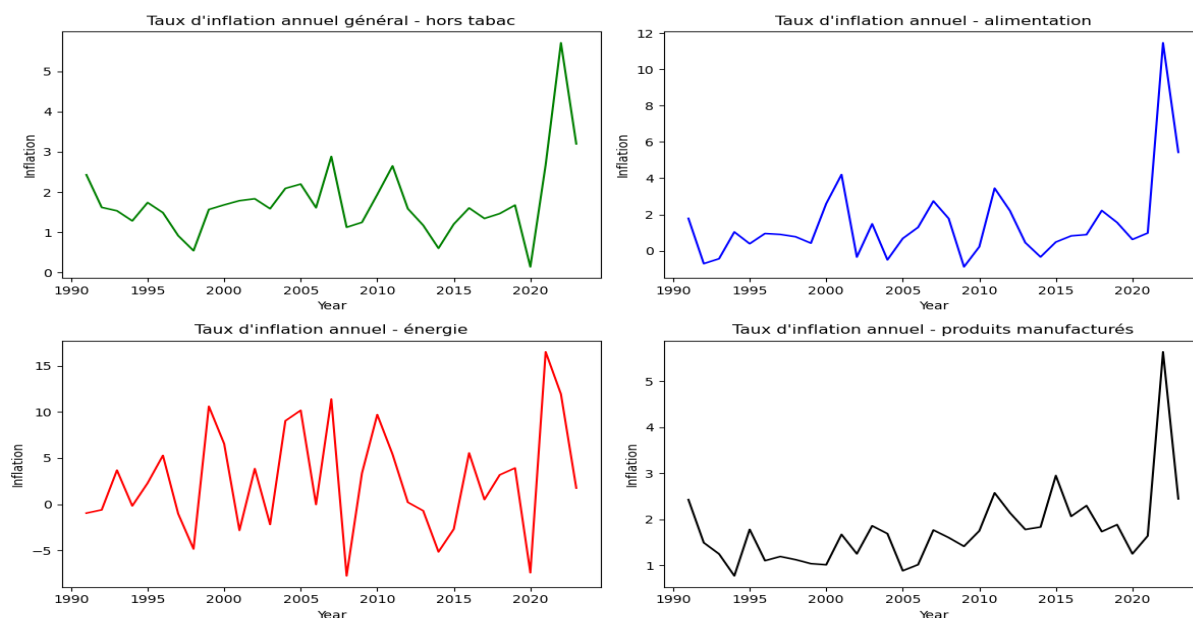


Figure 1 : Evolution du taux d'inflation annuel entre 1991 et 2024

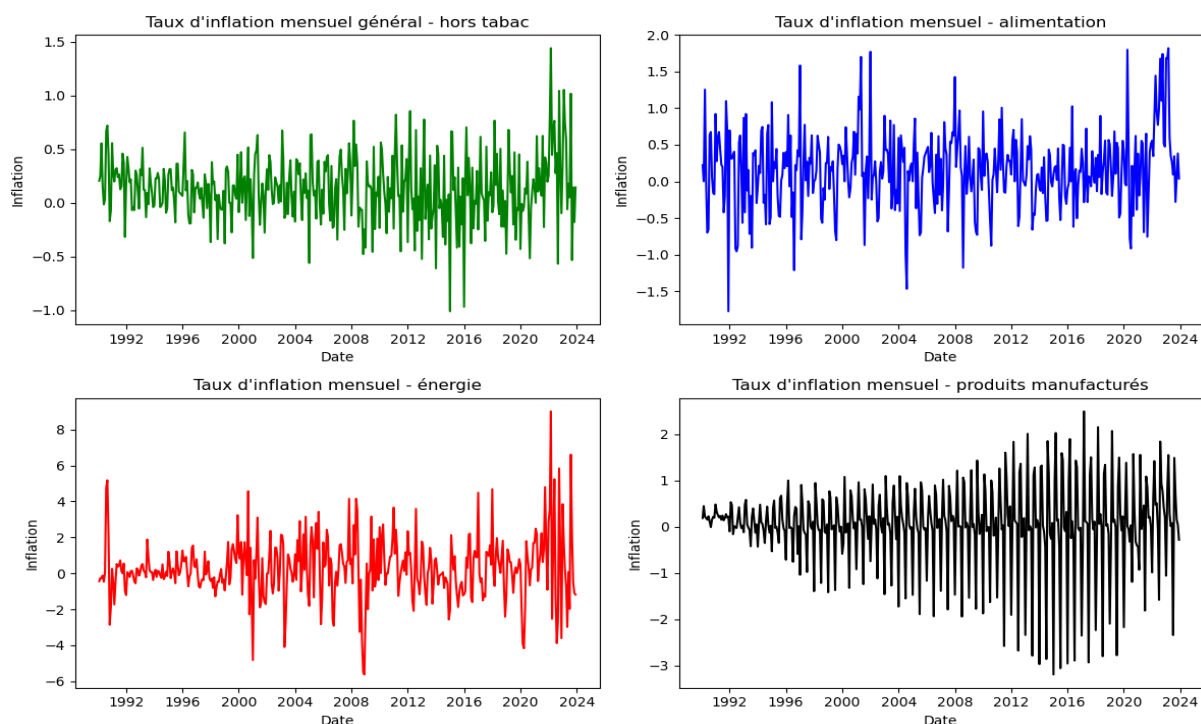


Figure 2 : Evolution du taux d'inflation mensuel entre 01/01/1990 et 01/12/2023

A partir de la figure 01 et 02, ces graphiques représentent l'évolution annuelle et mensuelle de l'inflation hors tabac, alimentaire, énergétique et produits manufacturés en France du 01/01/1990 au 01/12/2023. On peut constater que l'inflation fluctue considérablement au fil du temps, présentant une certaine volatilité et une série de pics et de creux. Par exemple, autour de 2022, l'inflation a grimpé d'une façon significative. Cette hausse s'explique principalement par trois facteurs principaux.

D'abord, la guerre en Ukraine a poussé à la hausse les prix de plusieurs matières premières, notamment le pétrole, le gaz, l'huile et le blé. Comme on peut voir sur les graphiques l'inflation énergétique est la plus élevée par rapport aux autres, plus de 15% en 2022 ce qui est énorme et peu souhaitable par les états car l'énergie représente une grande partie des budgets et donc une augmentation assez significative peut être reflétée et amortie que par la contribution des ménages et donc par une baisse de leur pouvoir d'achat.

Ensuite, cet événement a coïncidé avec la déclaration des pays de l'OPEP et leur décision de baisser leur production de pétrole et de gaz, un événement qui est venu renforcer ce phénomène et qui d'après nous a contribué à la hausse de l'inflation énergétique en France, surtout que le gazoduc North Stream s'est arrêté d'approvisionner l'Europe en gaz, et que les pays qui sont à proximité de la France sont sous contrat avec d'autres pays, comme l'Algérie et l'Italie, chose qui a poussé la France à s'approvisionner en gaz loin de ses frontières et plus spécialement au Qatar et aux États-Unis d'Amérique et au prix du marché.

Enfin, la dépréciation de l'euro a rendu les importations, particulièrement celles des énergies fossiles, plus coûteuses.

En 2020, l'inflation est tombée en dessous de 0 %, ce qui indique une désinflation due au ralentissement économique pendant la pandémie. Ces tendances nous aident à comprendre comment la politique et l'économie sont étroitement liées et que les événements mondiaux ont

une influence directe sur l'état d'un pays, sur sa stratégie commerciale internationale et sur sa stratégie monétaire.

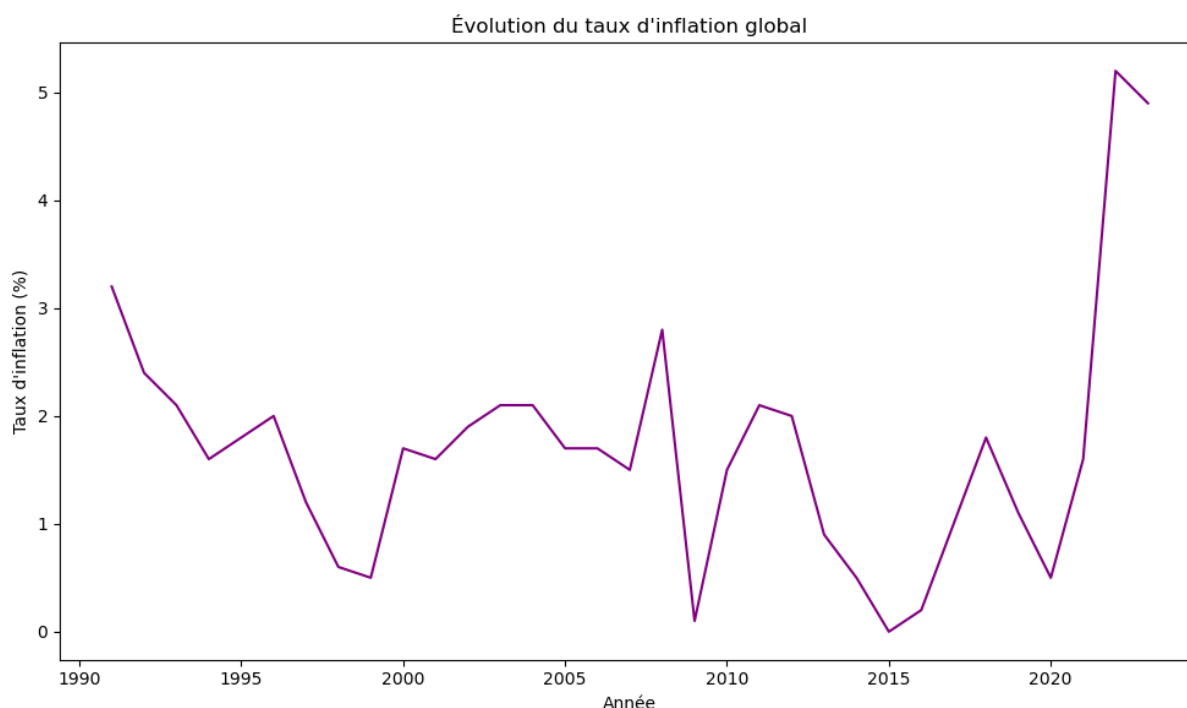


Figure 3 : Evolution du taux d'inflation général entre 1991 et 2023

Ce graphique représente l'évolution annuelle de l'inflation globale. On peut observer une tendance à la baisse entre 1990 et au début des années 2000 d'environ 3,2% à 0,5%. Ce déclin peut s'expliquer par plusieurs facteurs comme la récession du début des années 1990 avec une faible demande, les innovations technologiques par l'augmentation de la productivité et la réduction des coûts de production, l'Accord de libre-échange nord-américain (ALENA) en 1994 et l'entrée de la Chine dans l'Organisation mondiale du commerce (OMC) en 2001 ont contribué à la chute des prix et à l'éclatement de la bulle Internet en 2000.

On constate également une faible inflation entre 2011 et 2015 d'environ 2% à proche de 0%, influencée par les conséquences de la crise financière de 2008, la baisse des prix du pétrole entre 2014 et 2015, l'appréciation du dollar en rendant les importations moins chères et la diminution de la demande globale.

En 2021, l'inflation a considérablement augmenté jusqu'à atteindre 5% en 2023. Ce phénomène peut s'expliquer par plusieurs facteurs comme ceux cités précédemment, mais aussi par des facteurs tels que les mesures de relance post-COVID-19, les perturbations de la chaîne d'approvisionnement et les variations des prix des biens et de services.

Partie 02 : Analyse des données :

La question qu'on sait poser lors de la visualisation de nos données, on s'est dit que peut-être il y'aurait une relation de causalité entre la hausse de l'inflation énergétique et celle de l'inflation alimentaire et l'inflation des produits manufacturés. Notre raisonnement était que l'augmentation des prix de l'énergie pourrait avoir un effet sur l'augmentation des prix des produits alimentaires et des produits manufacturés, car ces deux industries avaient besoin de l'énergie pour pouvoir produire.

De fait, on a réalisé deux régressions linéaires pour essayer de déterminer une relation entre :

Mod1 : $Y_1 = a_1X + b + \text{erreur}$

Mod2 : $Y_2 = a_2X + b + \text{erreur}$

Avec :

- Y_1 = Taux d'inflation mensuel des produits alimentaires
- Y_2 = Taux d'inflation mensuel des produits manufacturés
- X = Taux d'inflation mensuel de l'énergie
- $b = 1$

OLS Regression Results						
=====						
Dep. Variable:	taux inflation mensuel alimentation	R-squared:	0.004			
Model:	OLS	Adj. R-squared:	0.002			
Method:	Least Squares	F-statistic:	1.735			
Date:	Tue, 16 Apr 2024	Prob (F-statistic):	0.189			
Time:	15:09:17	Log-Likelihood:	-312.78			
No. Observations:	407	AIC:	629.6			
Df Residuals:	405	BIC:	637.6			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

constante	0.1593	0.026	6.057	0.000	0.108	0.211
taux inflation mensuel énergie	0.0204	0.015	1.317	0.189	-0.010	0.051
=====						
Omnibus:	13.638	Durbin-Watson:	1.543			
Prob(Omnibus):	0.001	Jarque-Bera (JB):	27.032			
Skew:	0.118	Prob(JB):	1.35e-06			
Kurtosis:	4.240	Cond. No.	1.75			
=====						
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 4 : Résultat de la régression du modèle 01

OLS Regression Results						
Dep. Variable:	taux inflation mensuel produit manufacturés			R-squared:	0.002	
Model:	OLS			Adj. R-squared:	-0.001	
Method:	Least Squares			F-statistic:	0.7319	
Date:	Tue, 16 Apr 2024			Prob (F-statistic):	0.393	
Time:	15:09:12			Log-Likelihood:	-542.05	
No. Observations:	407			AIC:	1088.	
Df Residuals:	405			BIC:	1096.	
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
constante	0.0306	0.046	0.662	0.508	-0.060	0.121
taux inflation mensuel énergie	0.0233	0.027	0.855	0.393	-0.030	0.077
Omnibus:	79.888	Durbin-Watson:	2.243			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	160.484			
Skew:	-1.054	Prob(JB):	1.42e-35			
Kurtosis:	5.241	Cond. No.	1.75			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 5 : Résultat de la régression du modèle 02

Les résultats suggèrent que les deux modèles ne sont pas significatifs, on peut voir que les R^2 ajustés sont très bas, ce qui implique que le taux d'inflation mensuelle n'explique pas assez les deux phénomènes étudiés. On peut voir aussi que a_1 et a_2 sont positives mais très bas aussi ce qui implique qu'il y a une légère relation entre l'inflation mensuelle énergétique et celle alimentaire et des produits manufacturés mais qui reste statistiquement non significative.

Malgré qu'on sache très bien qu'il n'y pas de relation directe entre causalité et corrélation, on a voulu visualiser la corrélation entre ces variables pour démontrer la non significativité des deux modèles. (En utilisant la bibliothèque seaborn).

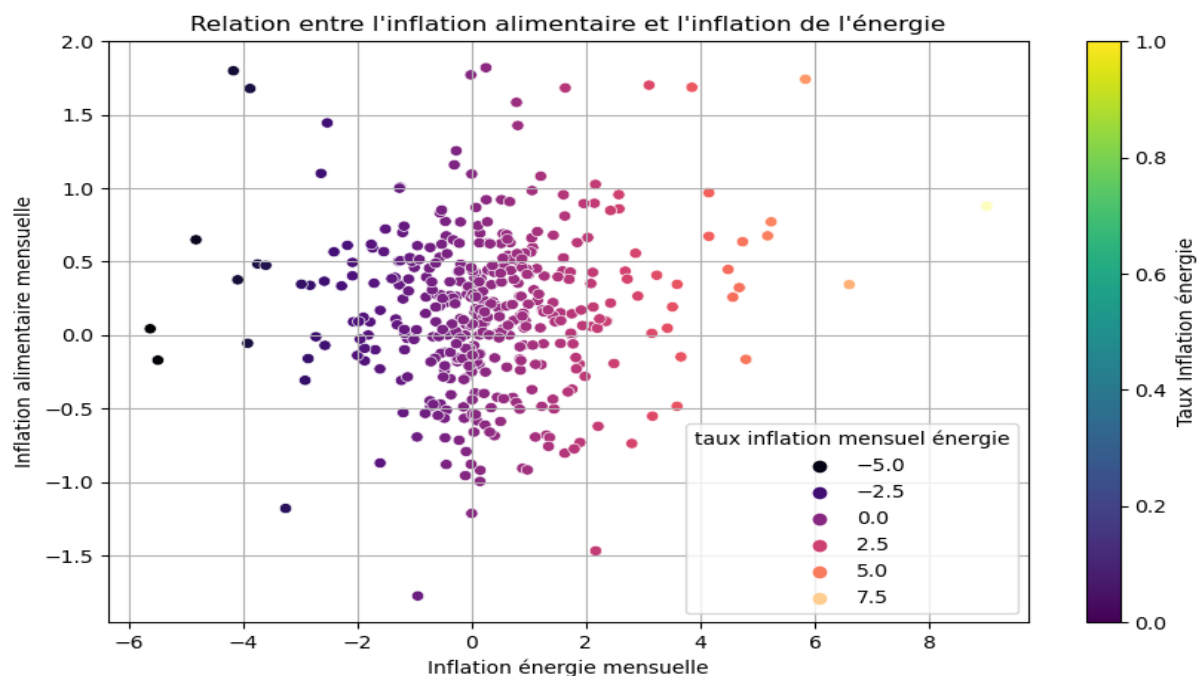


Figure 6 : Relation entre inflation alimentaire et inflation énergétique

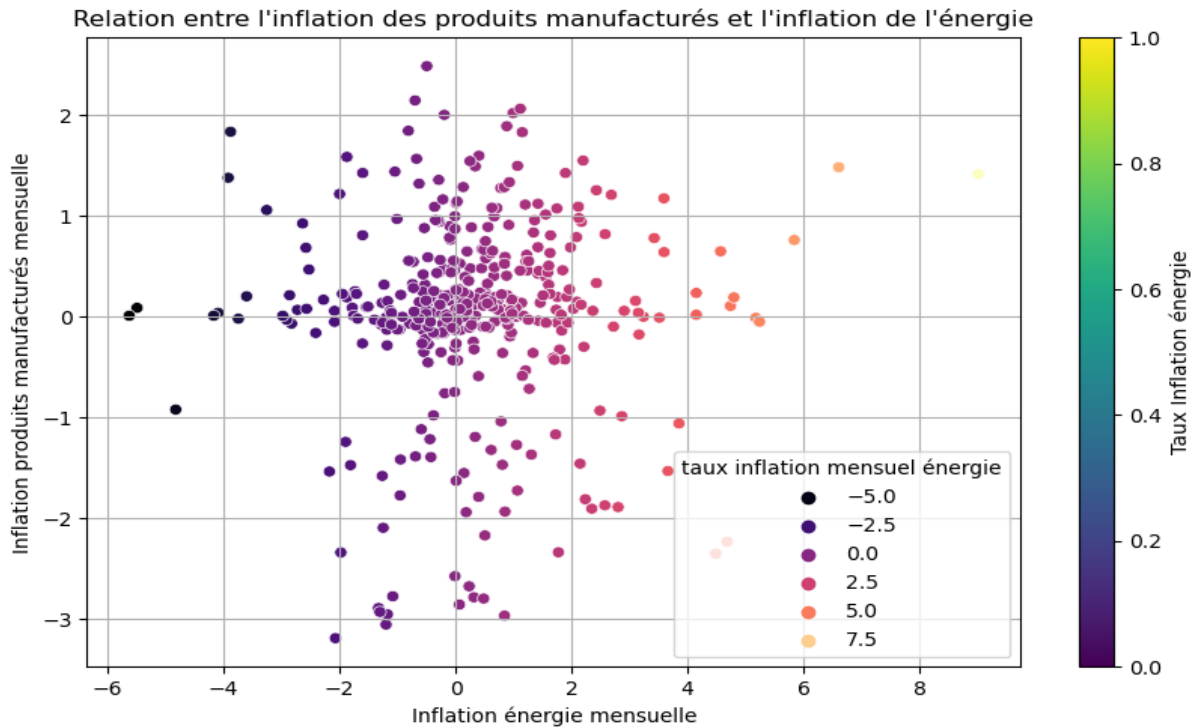


Figure 7 : Relation entre inflation alimentaire et inflation des produits manufacturés

D'après les graphiques et les calculs de corrélation qu'on a fait, on a trouvé que la corrélation entre l'inflation alimentaire et l'inflation énergétique est de 6,53% et la corrélation entre l'inflation des produits manufacturés et l'inflation énergétique est de 4,24%, relation positive et plus forte entre l'alimentation et l'énergie par rapport aux produits manufacturés et l'énergie mais qu'il ne sont pas significatifs statistiquement ce qui prouvent aussi la non significativité des deux modèles.

Ensuite, vu qu'il n'y avait pas de relation linéaire entre ces variables, on s'est dit que peut-être il y'aurait une relation non linéaire, c'est pour cela on a calculé la corrélation de Spearman, qui se calcul comme suit :

Équation 3 : Formule de calcul de la corrélation de Spearman

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

d_i : représente la variation entre les rangs d'observation dans les deux variable.

n : le nombre d'observation.

On a calculé aussi la corrélation de Kendall, dont le calcul se déroule de cette manière :

Équation 4 : Formule de calcul de la corrélation de Kendall

$$\tau = \frac{\text{Number of concordant pairs} - \text{Number of discordant pairs}}{n(n-1)/2}$$

n : le nombre de paires.

Ces taux de corrélation sont tous les deux non paramétrique (c'est-à-dire non basé sur des distributions sous-jacentes des données comme la loi normale) et on est tombé sur le même résultat avec un taux de corrélation très bas eu un p-value supérieure à 5%.

Et donc selon notre jeu de donnée, il n'y a pas de relation de causalité entre l'inflation énergétique et l'inflation alimentaire et celle des produits manufacturés.

Maintenant, on a voulu voir l'impact de l'inflation mensuelle alimentaire, énergétique et celle des produits manufacturés sur l'inflation mensuelle globale hors tabac. De fait, on a réalisé une régression multiple en utilisant la bibliothèque statsmodels. Notre modèle est le suivant :

$$\text{Mod3 : } Y = a_1X_1 + a_2X_2 + a_3X_3 + b + \text{erreur}$$

Avec :

- Y : Taux d'inflation mensuel hors tabac
- X_1 : Taux d'inflation mensuel de l'énergie
- X_2 : Taux d'inflation mensuel des produits alimentaires
- X_3 : Taux d'inflation mensuel des produits manufacturés
- $b = 1$

OLS Regression Results

Dep. Variable:

taux inflation mensuel hors tabac

R-squared:

0.668

Model:

OLS

Adj. R-squared:

0.665

Method:

Least Squares

F-statistic:

269.9

Date:

Tue, 16 Apr 2024

Prob (F-statistic):

5.20e-96

Time:

17:01:09

Log-Likelihood:

126.01

No. Observations:

407

AIC:

-244.0

Df Residuals:

403

BIC:

-228.0

Df Model:

3

Covariance Type:

nonrobust

	coef	std err	t	P> t	[0.025	0.975]
const	0.0747	0.009	7.973	0.000	0.056	0.093
taux inflation mensuel énergie	0.0811	0.005	15.292	0.000	0.071	0.092
taux inflation mensuel alimentation	0.1795	0.017	10.584	0.000	0.146	0.213
taux inflation mensuel produit manufacturés	0.1925	0.010	19.943	0.000	0.174	0.212

Omnibus:

152.044

Durbin-Watson:

2.123

Prob(Omnibus):

0.000

Jarque-Bera (JB):

680.153

Skew:

-1.582

Prob(JB):

2.03e-148

Kurtosis:

8.486

Cond. No.

3.34

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 8 : Résultat régression multiple du modèle 3

Les résultats suggèrent que notre modèle est significatif, on peut voir que le R^2 ajustés est égale à 66,8%, ce qui implique que nos trois variables indépendantes expliquent assez bien la variance de notre variable expliquée. On peut voir aussi que a_1 , a_2 et a_3 sont positives, ce qui démontre une relation positive entre nos variables. On remarque aussi que la F statistique est assez élevée avec une probabilité inférieure à 5%, ce qui confirme la significativité statistique de notre modèle 3. Les corrélations entre les variables indépendantes et la variable expliquée :

taux inflation mensuel alimentation	0.354614
taux inflation mensuel énergie	0.484693
taux inflation mensuel produit manufacturés	0.603466

Figure 9 : Corrélation entre les variables explicatives et la variable expliquée

On remarque que les corrélations sont positives c'est-à-dire qu'il y a une relation positive entre les variables indépendantes la variable dépendante (ils évoluent dans la même direction), la corrélation entre l'inflation générale hors tabac et l'inflation des biens manufacturés est plus élevée par rapport aux autres variables indépendantes.

Partie 03 : Application du modèle ARIMA :

Data frame utilisé : df_5 (taux d'inflation générale entre 1991 et 2023)

Vu qu'on travaille sur l'inflation, on sait très bien que l'inflation est une série temporelle dite non stationnaire (car elle présente des tendances qui changent au fur et à mesure que le temps passe (des pics et des creux), elle présente aussi une saisonnalité remarquable et une variance non constante). La figure 10 vient confirmer les deux premières conditions de la non stationnarité de la série temporelle de l'inflation annuelle globale, qui sont la variation des tendances et la présence d'une saisonnalité. La figure 11 vient confirmer la troisième condition, qui est la variation de la variance de la série. La volatilité de la série peut être expliquée par les différents risques présents sur la scène internationale et leurs formes qui sont imprévisibles.



Figure 10 : Représentation graphique de la décomposition saisonnière de notre série temporelle df_5

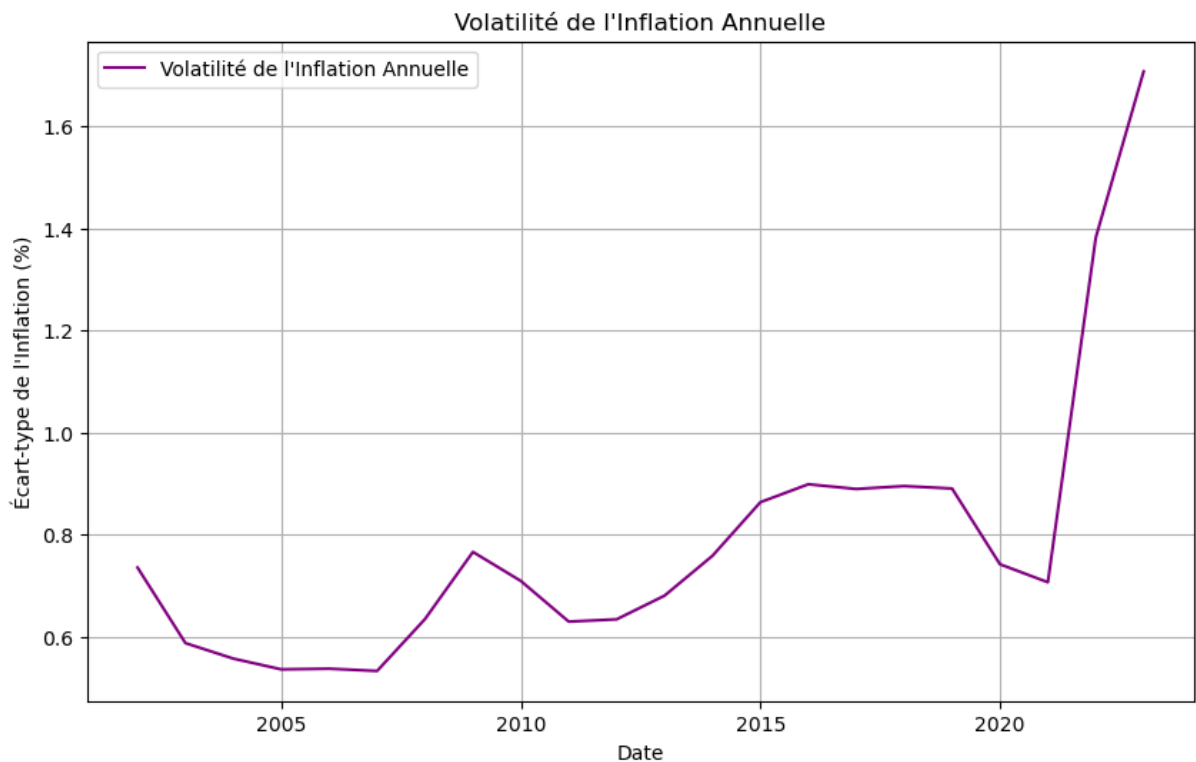


Figure 11 : Représentation graphique de la volatilité de l'inflation annuelle

Et pour cela, on a décidé d'utiliser le modèle ARIMA, qui est un modèle d'analyse et de prédiction statistique basé sur l'apprentissage (le Machine Learning), Ce modèle viendra corriger notre série et la rendre stationnaire avant d'appliquer le modèle ARMA afin de la modéliser et de prédire le taux d'inflation annuelle pour l'année 2024.

1- Analyse de la série temporelle :

Dans cette partie, on a décidé d'analyser l'évolution de l'inflation annuelle globale par rapport au modèle ARIMA, pour cela on prit 60% des données de df_5 pour entraîner le modèle et 40% pour le tester. Le résultat est le suivant :

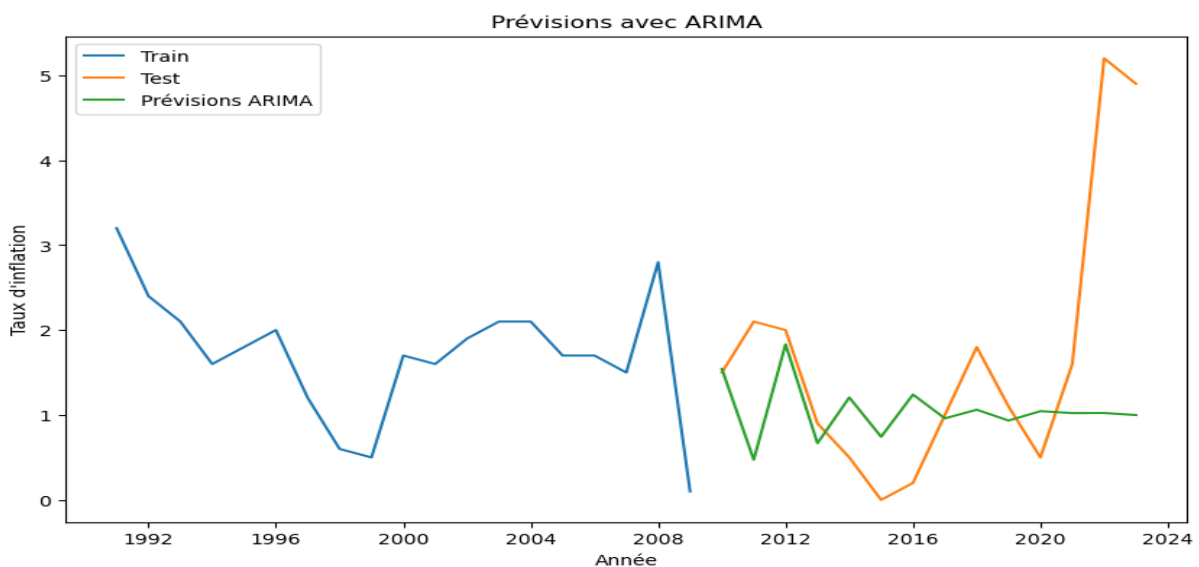


Figure 12 : Représentation graphique de la série temporelle df_5 (application modèle ARIMA)

Ensuite, on a calculé l'erreur quadratique moyenne (RMSE) qui représente un indicateur de précision de la prédiction des modèles statistique, il se calcul comme suit :

Équation 5 : Formule de calcul de l'erreur quadratique moyenne

$$EQM = \frac{\sum_{i=0}^n (\hat{y}_i - y_i)^2}{n}$$

\hat{y}_i : les valeurs prédites par le modèle.
 Y_i : les valeurs réelles.
 N : le nombre d'observation.

Ce qu'on remarque est qu'il y a une différence flagrante entre les valeurs réelles de l'inflation et les valeurs prédites par le modèle, il y a plusieurs explications possibles à ce résultat.

D'abord, le nombre d'observation qui est insuffisant pour avoir une estimation plus ajustée. En plus, le modèle ne prend en compte qu'une seule variable explicative.

Ensuite, l'inflation est reliée à des événements imprévisibles et des politiques économiques qui peuvent varier et changer d'année en année. Prenons l'exemple de 2022, le modèle prédit que le taux d'inflation soit d'environ 1%, mais réellement l'inflation a dépassé les 5% à cause de la guerre en Ukraine, événement que aucun modèle peut prévoir.

Enfin, l'erreur quadratique moyenne est égale à 1,66 qui assez élevé, qui indique que notre modèle appliqué sur l'inflation manque de précision.

2- Prédiction pour 2024 :

Afin de prédire le taux d'inflation en France pour l'année 2024, on a appliqué la même méthode d'apprentissage sauf que on a utilisé toutes les données pour entrainer le modèle (training dataset = 100%).

Le modèle ARIMA a prédit que le taux d'inflation 2024 sera d'environ 3,54%, résultat qu'on trouve disons plus cohérent que les résultats précédents, vu qu'on a augmenté les données d'entraînement, mais malgré ça, on trouve qu'il reste assez loin de la réalité vu la conjoncture internationale actuelle, on peut citer : la guerre à Gaza, l'attaque de l'Iran sur Israël, la fermeture du Golfe d'Aden par les Houthis et la flambé des couts de transport.

On estime que le taux d'inflation va avoir probablement une tendance haussière ou une stagnation malgré les politiques mises en place par le gouvernement.

Partie 04 : Comparaison des indices de prix à la consommation harmonisé de plusieurs pays entre janvier 2019 et décembre 2023 :

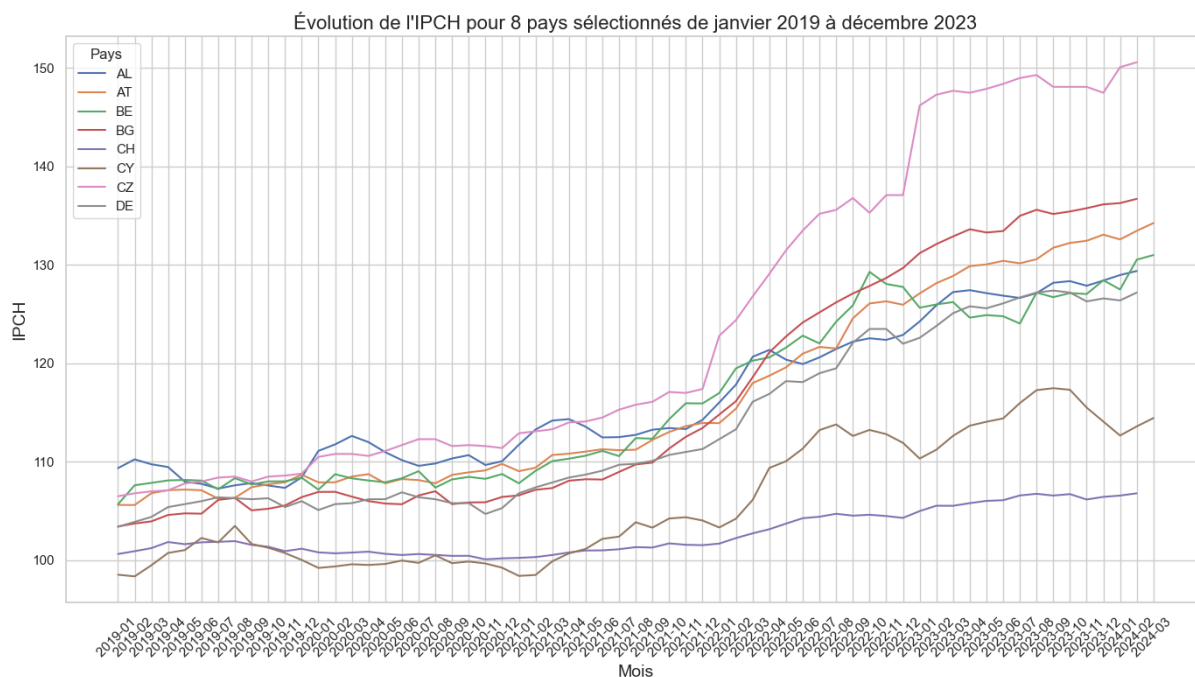


Figure 13 : Evolution de l'IPCH entre Jan.2019 et Dec.2023

Ce graphique montre l'évolution de l'indice des prix à la consommation harmonisé (IPCH) de janvier 2019 à décembre 2023 pour huit pays européens, qui sont l'Albanie (AL, en bleu), l'Autriche (AT, en orange), la Belgique (BE, en vert), la Bulgarie (BG, en rouge), la Suisse (CH, en violet), le Chypre (CY, en marron), la Tchèque (CZ, en rose), et l'Allemagne (DE, en gris). Nous observons une tendance à la hausse de l'IPCH pour tous les pays, ce qui indique que le niveau général des prix a augmenté au cours de cette période. Nous observons aussi des évolutions similaires pour la Bulgarie, l'Autriche, l'Albanie, l'Allemagne, et la Belgique. La Suisse montre une courbe relativement plate comparé aux autres pays, indiquant une inflation stable et contrôlée. Cela peut être dû à sa politique monétaire et au fait qu'elle n'est pas membre de l'Union européenne. En fin 2023, la Tchèque a connu des pics d'inflation importants, principalement influencés par la guerre en Ukraine, la baisse des salaires et la crise énergétique.

Conclusion :

L'inflation est un phénomène complexe et crucial pour un état, pour cette raison il faut une gestion rigoureuse et visionnaire pour essayer d'anticiper toute crise.

Les banques centrales de la zone euro opèrent une forte surveillance sur l'inflation, en mettent en œuvre des politiques monétaires, en ajustant les taux d'intérêts, et en maintenant un taux d'inflation à 2%. L'inflation reste très subtile car il est difficile de déterminer la principale raison des fluctuations des prix, car d'autres facteurs tels que les variations des coûts, les chocs de l'offre et de la demande, la quantité de monnaie en circulation, les changements politiques, les crises économiques, les conditions de concurrence jouent tous un rôle dans la hausse ou la baisse des prix.

Les graphiques que nous avons créés avec les données nous permettent de comprendre les liens entre les changements économiques et d'identifier les facteurs dans les tendances générales. Néanmoins, l'analyse de l'inflation qu'on a réalisée reste très limitée, ce qui nécessite des données supplémentaires pour approfondir nos réflexions et d'améliorer nos résultats.

Annexe :

```
#Importation des données
Directory_path = r"C:\Users\DELL\OneDrive\Bureau\py2"
df_1 = pd.read_csv(os.path.join(Directory_path, "Indice_prix_consommation_90_24.csv"), sep = ";") #Ce fichier contient
df_2 = pd.read_csv(os.path.join(Directory_path, "Indice_prix_consommation_90_24_alimentation.csv"), sep = ";") #Ce fic
df_3 = pd.read_csv(os.path.join(Directory_path, "Indice_prix_consommation_90_24_energie.csv"), sep = ";") #Ce fichier
df_4 = pd.read_csv(os.path.join(Directory_path, "Indice_prix_consommation_90_24_produit_manufacturé.csv"), sep = ";")
df_5 = pd.read_excel(os.path.join(Directory_path, "Taux d'inflation annuel.xlsx")) #Ce fichier contient les taux d'inf
```

Figure 14 : Importation des données

```
#Dans chaque df, on va garder les colonnes Date et indice et on va supprimer le reste. On va supprimer aussi les 3 premières lignes
#Ensuite on renommer les colonnes
df_1.drop(columns = ["Codes", "Unnamed: 3"], axis = 1, inplace = True)
df_2.drop(columns = ["Codes"], axis = 1, inplace = True)
df_3.drop(columns = ["Codes"], axis = 1, inplace = True)
df_4.drop(columns = ["Codes"], axis = 1, inplace = True)
df_1.drop([0, 1, 2], inplace = True)
df_2.drop([0, 1, 2], inplace = True)
df_3.drop([0, 1, 2], inplace = True)
df_4.drop([0, 1, 2], inplace = True)
df_5.drop([33, 34, 35, 36], inplace = True)
#On va réinitialiser les index
df_1.reset_index(drop = True, inplace = True)
df_2.reset_index(drop = True, inplace = True)
df_3.reset_index(drop = True, inplace = True)
df_4.reset_index(drop = True, inplace = True)
#On va renommer les colonnes
df_1.rename(columns = {"Libellé" : "Date", "Indice des prix à la consommation - Base 2015 - Ensemble des ménages - France - Ensemble hors tabac" : "indice hors tabac"}, inplace = True)
df_2.rename(columns = {"Libellé" : "Date", "Indice des prix à la consommation - Base 2015 - Ensemble des ménages - France - Alimentation" : "indice alimentation"}, inplace = True)
df_3.rename(columns = {"Libellé" : "Date", "Indice des prix à la consommation - Base 2015 - Ensemble des ménages - France - Énergie" : "indice énergie"}, inplace = True)
df_4.rename(columns = {"Libellé" : "Date", "Indice des prix à la consommation - Base 2015 - Ensemble des ménages - France - Produits manufacturés" : "indice produit manufacturés"}, inplace = True)
df_5.rename(columns = {"Année" : "date", "Inflation" : "taux d'inflation"}, inplace = True)
df_5.info()
```

Figure 15 : Nettoyage des données

```
#A partir de ces 4 bases, on va créer la base IPC_df
IPC_df = pd.merge(df_1, df_2, on = "Date", how = "inner")
IPC_df = pd.merge(IPC_df, df_3, on = "Date", how = "inner")
IPC_df = pd.merge(IPC_df, df_4, on = "Date", how = "inner")
display(IPC_df)
IPC_df.info()
```

Figure 16 : Fusion des données

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 411 entries, 0 to 410
Data columns (total 5 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Date                                411 non-null    object
1   indice hors tabac                    411 non-null    object
2   indice alimentation                  411 non-null    object
3   indice énergie                       411 non-null    object
4   indice produit manufacturés          411 non-null    object
dtypes: object(5)
memory usage: 16.2+ KB
```

Figure 17 : Information de la data frame "IPC_df"

```
#
df_5["Année"] = pd.to_datetime(df_5["date"], format = "%Y")
# maintenant on va convertir la colonne Date en type Datetime et les colonnes indices en type numérique (c-à-d : float)
IPC_df["Date"] = pd.to_datetime(IPC_df["Date"], format = "%Y-%m")
IPC_df["indice alimentation"] = pd.to_numeric(IPC_df["indice alimentation"], errors = "coerce")
IPC_df["indice hors tabac"] = pd.to_numeric(IPC_df["indice hors tabac"], errors = "coerce")
IPC_df["indice énergie"] = pd.to_numeric(IPC_df["indice énergie"], errors = "coerce")
IPC_df["indice produit manufacturés"] = pd.to_numeric(IPC_df["indice produit manufacturés"], errors = "coerce")
IPC_df.info()
```

Figure 18 : Conversion des variables

```
#Création d'une nouvelle base de données nommée IPC_df_annuelle qui contient les taux d'inflation annuel pour chaque indice
IPC_df['Year'] = IPC_df['Date'].dt.year
IPC_df_annuelle = IPC_df.groupby('Year').agg({
    'indice hors tabac': lambda x: (x.iloc[-1] - x.iloc[0]) / x.iloc[0] * 100,
    'indice alimentation': lambda x: (x.iloc[-1] - x.iloc[0]) / x.iloc[0] * 100,
    'indice énergie': lambda x: (x.iloc[-1] - x.iloc[0]) / x.iloc[0] * 100,
    'indice produit manufacturés': lambda x: (x.iloc[-1] - x.iloc[0]) / x.iloc[0] * 100
}).rename(columns={
    'indice hors tabac': 'Taux d\'inflation hors tabac',
    'indice alimentation': 'Taux d\'inflation alimentation',
    'indice énergie': 'Taux d\'inflation énergie',
    'indice produit manufacturés': 'Taux d\'inflation produits manufacturés'
})
if IPC_df_annuelle.index[0] == 1990:
    IPC_df_annuelle = IPC_df_annuelle.drop(IPC_df_annuelle.index[0])
IPC_df_annuelle.reset_index(inplace=True)
display(IPC_df_annuelle)
IPC_df_annuelle.info()
IPC_df_annuelle.describe()
```

Figure 19 : Création de "IPC_df_annuelle"

```
#Création d'une nouvelle base de données nommée IPC_df_mensuelle qui contient les taux d'inflation mensuelle pour chaque indice
IPC_df = IPC_df.sort_values(by = "Date")
IPC_df["taux inflation mensuel hors tabac"] = IPC_df["indice hors tabac"].pct_change() * 100
IPC_df["taux inflation mensuel alimentation"] = IPC_df["indice alimentation"].pct_change() * 100
IPC_df["taux inflation mensuel énergie"] = IPC_df["indice énergie"].pct_change() * 100
IPC_df["taux inflation mensuel produit manufacturés"] = IPC_df["indice produit manufacturés"].pct_change() * 100
IPC_df_mensuelle = IPC_df[["Date", "taux inflation mensuel hors tabac", "taux inflation mensuel alimentation", "taux inflation mensuel énergie", "taux inflation mensuel produit manufacturés"]]
IPC_df_mensuelle.drop([0], inplace = True)
display(IPC_df_mensuelle)
IPC_df_mensuelle.info()
IPC_df_mensuelle.describe()
```

Figure 20 : Création de "IPC_df_mensuelle"

Sitographie :

- [Indice des prix à la consommation - Base 2015 - Ensemble des ménages - France - Ensemble hors tabac | Insee](#)
- [Indice des prix à la consommation - Base 2015 - Ensemble des ménages - France - Alimentation | Insee](#)
- [Indice des prix à la consommation - Base 2015 - Ensemble des ménages - France - Énergie | Insee](#)
- [Indice des prix à la consommation - Base 2015 - Ensemble des ménages - France - Produits manufacturés | Insee](#)
- [Taux d'inflation | Insee](#)
- https://ec.europa.eu/eurostat/databrowser/view/prc_hicp_midx/default/table?lang=en&category=prc.prc_hicp

Table des illustrations :

Figure 1 : Evolution du taux d'inflation annuel entre 1991 et 2024	5
Figure 2 : Evolution du taux d'inflation mensuel entre 01/01/1990 et 01/12/2023	6
Figure 3 : Evolution du taux d'inflation général entre 1991 et 2023	7
Figure 4 : Résultat de la régression du modèle 01	8
Figure 5 : Résultat de la régression du modèle 02	9
Figure 6 : Relation entre inflation alimentaire et inflation énergétique	9
Figure 7 : Relation entre inflation alimentaire et inflation des produits manufacturés	10
Figure 8 : Résultat régression multiple du modèle 3	11
Figure 9 : Corrélation entre les variables explicatives et la variable expliquée	11
Figure 10 : Représentation graphique de la décomposition saisonnière de notre série temporelle df_5	12
Figure 11 : Représentation graphique de la volatilité de l'inflation annuelle	13
Figure 12 : Représentation graphique de la série temporelle df_5 (application modèle ARIMA)	13
Figure 13 : Evolution de l'IPCH entre Jan.2019 et Dec.2023	15
Figure 14 : Importation des données	17
Figure 15 : Nettoyage des données	17
Figure 16 : Fusion des données	17
Figure 17 : Information de la data frame "IPC_df"	17
Figure 18 : Conversion des variables	17
Figure 19 : Création de "IPC_df_annuelle"	18
Figure 20 : Création de "IPC_df_mensuelle"	18
 Tableau 1 : Liste des bibliothèques utilisées	 2
Tableau 2 : IPC_df DATA	3
Tableau 3 : IPC_df_annuelle DATA	4
Tableau 4 : IPC_df_mensuelle DATA	5
 Équation 1 : Formule de calcul du taux d'inflation annuel	 4
Équation 2 : Formule de calcul du taux d'inflation mensuel	5
Équation 3 : Formule de calcul de la corrélation de Spearman	10
Équation 4 : Formule de calcul de la corrélation de Kendall	10
Équation 5 : Formule de calcul de l'erreur quadratique moyenne	14