

Rapport Projet : MODELES DE REGRESSION

Introduction :

Membres du groupe :

Domaine :

Problématique :

Besoins d'analyse :

Source des Données :

Dictionnaire des variables et des modalités :

Préparation des données :

DB_regression.xlsx :

Prétraitement :

Modélisation statistique :

Statistiques descriptives :

Résumé statistique :

Distributions :

Evolution en fonction du temps :

Boxplot :

Analyse bivariée :

Heatmap :

Feature "X" vs Fertility Rate :

Scatter plot : Feature "X" vs Fertility Rate

Présélection des variables à fortes corrélation pour la regression :

RLS :

RLS avec les attributs sélectionnées :

Visualisation :

Sélection manuelle :

Cross validation :

Régularisation :

Lasso :

Ridge :

RLM:

Extra : **Random forest regressor + XGBoost**

Random Forest regressor

XGBoost:

Conclusion :

Introduction :

Membres du groupe :

Nom	Filière	Numéro
EL HASNAOUI Chaymae	ISSIC	8
TOUMZITE Ali	ICSD	42
LAMINE Mohamed Amine	ICSD	30

Domaine :

La démographie

Problématique :

Effectuer des prédictions démographiques

Besoins d'analyse :


Les études démographiques sont essentielles pour comprendre les tendances de la population, les mouvements migratoires, les taux de natalité et de mortalité, ainsi que d'autres aspects cruciaux de la composition démographique d'une région ou d'un pays. Pour garantir une planification efficace des politiques publiques, des programmes sociaux, et des stratégies de développement, une analyse approfondie des données démographiques s'avère indispensable. Ce document expose les besoins essentiels en matière d'analyse pour mener à bien des études démographiques significatives.

L'objectif principal de cette analyse est de fournir une compréhension approfondie de la dynamique démographique d'une population spécifique, en mettant en évidence les facteurs clés influençant sa structure et sa croissance.

Source des Données :

DataBank | The World Bank

DataBank is an analysis and visualisation tool that contains collections of time series data on a variety of topics where you can create your own queries, generate tables, charts and maps and easily save, embed and share them

 <https://databank.worldbank.org/home>

Dictionnaire des variables et des modalités :

Indicator Name	Topic	Long definition	Range of values
----------------	-------	-----------------	-----------------

Age dependency ratio (% of working-age population)	Health: Population: Dynamics	Age dependency ratio is the ratio of dependents--people younger than 15 or older than 64--to the working-age population--those ages 15-64. Data are shown as the proportion of dependents per 100 working-age population.	40% - 60%
Age dependency ratio, old (% of working-age population)	Health: Population: Dynamics	Age dependency ratio, old, is the ratio of older dependents--people older than 64--to the working-age population--those ages 15-64. Data are shown as the proportion of dependents per 100 working-age population.	20% - 30%
Age dependency ratio, young (% of working-age population)	Health: Population: Dynamics	Age dependency ratio, young, is the ratio of younger dependents--people younger than 15--to the working-age population--those ages 15-64. Data are shown as the proportion of dependents per 100 working-age population.	20% - 30%
Birth rate, crude (per 1,000 people)	Health: Population: Dynamics	Crude birth rate indicates the number of live births occurring during the year, per 1,000 population estimated at midyear. Subtracting the crude death rate from the crude birth rate provides the rate of natural increase, which is equal to the rate of population change in the absence of migration.	10 - 30
Death rate, crude (per 1,000 people)	Health: Population: Dynamics	Crude death rate indicates the number of deaths occurring during the year, per 1,000 population estimated at midyear. Subtracting the crude death rate from the crude birth rate provides the rate of natural increase, which is equal to the rate of population change in the absence of migration.	5 - 15
Fertility rate, total (births per woman)	Health: Reproductive health	Total fertility rate represents the number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with age-specific fertility rates of the specified year.	1.5 - 3.5
Life expectancy at birth, total (years)	Health: Mortality	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of	70 - 85

		mortality at the time of its birth were to stay the same throughout its life.	
Net migration	Social Protection & Labor: Migration	Net migration is the net total of migrants during the period, that is, the number of immigrants minus the number of emigrants, including both citizens and noncitizens.	-50,000 to 50,000 (varies widely by country)
Urban population	Environment: Density & urbanization	Urban population refers to people living in urban areas as defined by national statistical offices. It is calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects. Aggregation of urban and rural population may not add up to total population because of different country coverages.	Varies widely depending on country and region
Urban population growth (annual %)	Environment: Density & urbanization	Urban population refers to people living in urban areas as defined by national statistical offices. It is calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects.	0.5% - 3%
Rural population growth (annual %)	Environment: Density & urbanization	Rural population refers to people living in rural areas as defined by national statistical offices. It is calculated as the difference between total population and urban population.	-2% to 2%
Rural population	Environment: Density & urbanization	Rural population refers to people living in rural areas as defined by national statistical offices. It is calculated as the difference between total population and urban population. Aggregation of urban and rural population may not add up to total population because of different country coverages.	Varies widely depending on country and region
Population, total	Health: Population: Structure	Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. The values shown are midyear estimates.	Varies widely depending on country and region
Population growth (annual %)	Health: Population: Dynamics	Annual population growth rate for year t is the exponential rate of growth of	Varies widely depending on

		midyear population from year t-1 to t, expressed as a percentage . Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship.	country and region
Population ages 65 and above, total	Health: Population: Structure	Total population 65 years of age or older. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship.	Varies widely depending on country and region
Population ages 15-64, total	Health: Population: Structure	Total population between the ages 15 to 64. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship.	Varies widely depending on country and region
Mortality rate, infant (per 1,000 live births)	Health: Mortality	Infant mortality rate is the number of infants dying before reaching one year of age, per 1,000 live births in a given year.	1 - 50

Préparation des données :

DB_regression.xlsx :

DB_regression.xlsx

esiac-my.sharepoint.com/cx/r/personal/mohamed-amine_lamine_esi_ac_ma/_layouts/15/Doc.aspx

DB_regression

Search for tools, help, and more (Alt + Q)

FileHomeInsertSharePage LayoutFormulasDataReviewViewAutomateHelpDraw

CommentsCatch upEditingShare

Calibri (Body)11

General

\$%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

10

Prétraitement :

- Pas de valeurs manquantes
- Toutes les valeurs sont converties en des valeurs numériques
- 63 observations
- 17 indicateurs démographique
- Changement de nomenclature long en une plus adéquate et significatif
- variable cible : 'Fertility Rate '

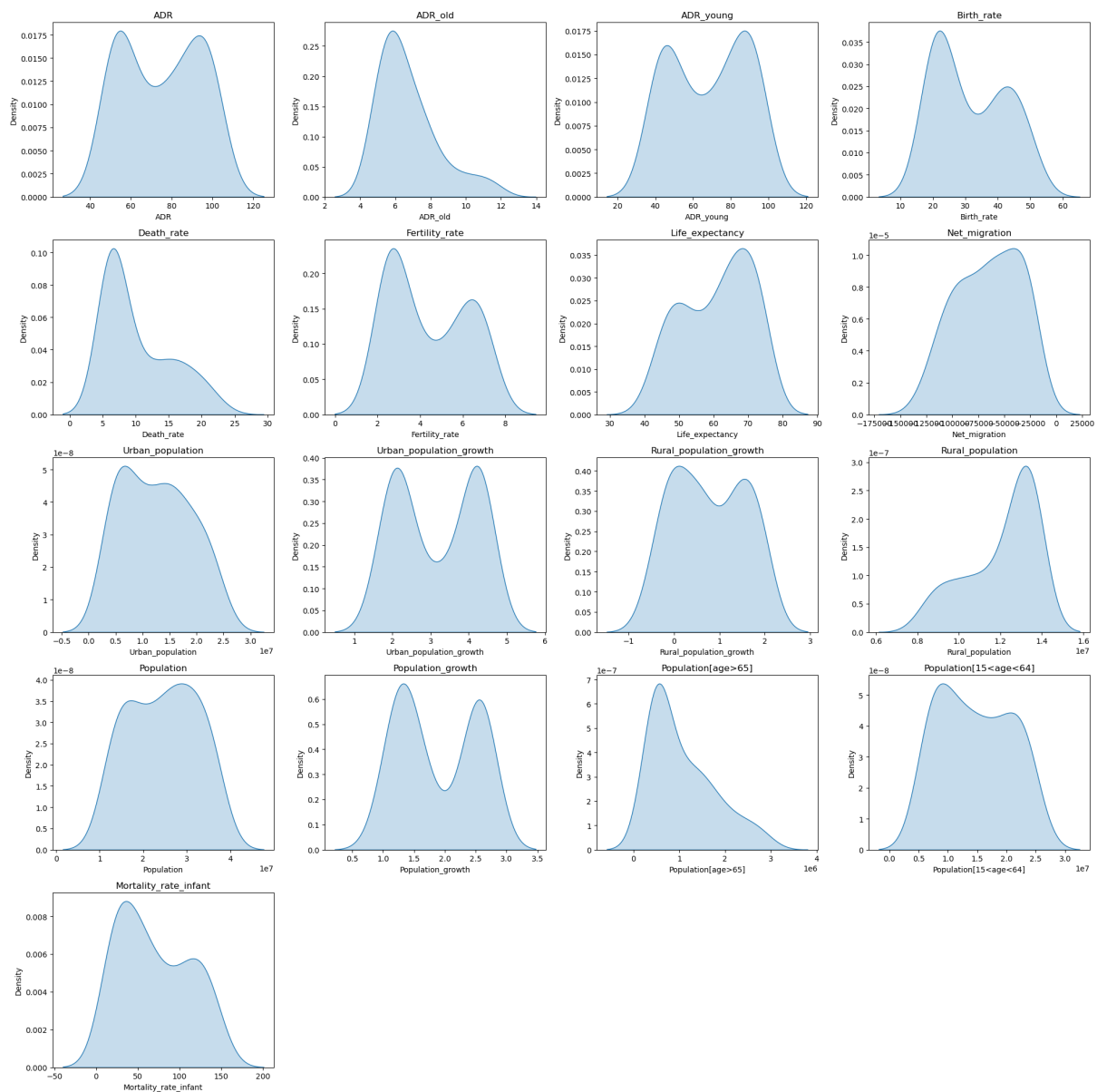
Modélisation statistique :

Statistiques descriptives :

Résumé statistique :

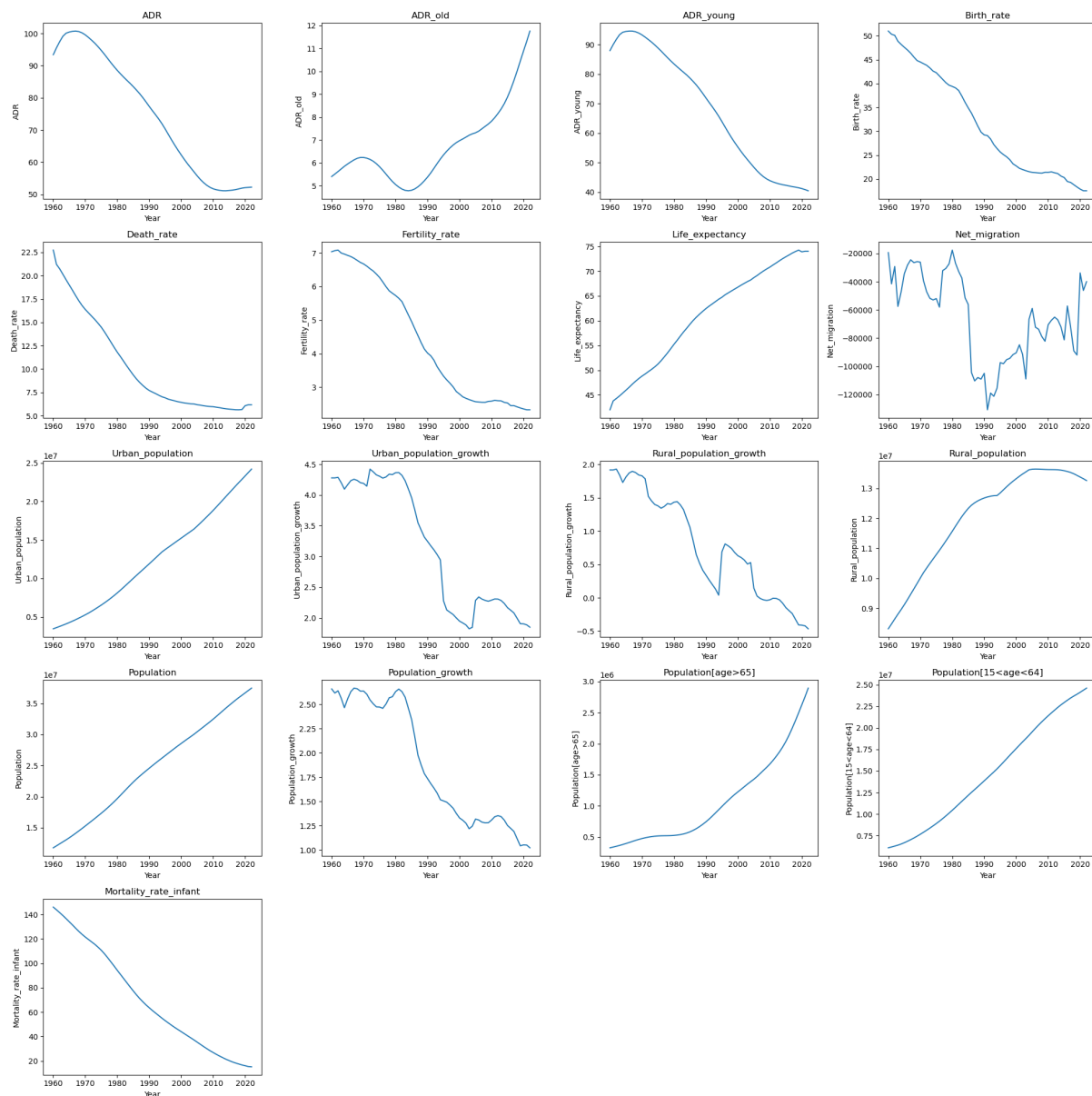
	count	mean	std	min	25%	50%	75%	max
ADR	63.0	7.494662e+01	1.869572e+01	5.110523e+01	5.411814e+01	7.598363e+01	9.344559e+01	1.007218e+02
ADR_old	63.0	6.732500e+00	1.688843e+00	4.781439e+00	5.534682e+00	6.197075e+00	7.435284e+00	1.175197e+01
ADR_young	63.0	6.821412e+01	1.999329e+01	4.046611e+01	4.668286e+01	7.042407e+01	8.790663e+01	9.460131e+01
Birth_rate	63.0	3.158754e+01	1.086598e+01	1.754500e+01	2.140650e+01	2.909200e+01	4.191400e+01	5.097500e+01
Death_rate	63.0	1.023521e+01	5.091669e+00	5.637000e+00	6.170000e+00	7.530000e+00	1.423900e+01	2.274500e+01
Fertility_rate	63.0	4.365317e+00	1.785232e+00	2.328000e+00	2.596000e+00	3.937000e+00	6.205000e+00	7.088000e+00
Life_expectancy	63.0	6.087797e+01	9.906258e+00	4.201700e+01	5.159000e+01	6.294600e+01	6.936600e+01	7.427000e+01
Net_migration	63.0	-6.565043e+04	3.063638e+04	-1.307430e+05	-9.165450e+04	-6.513600e+04	-3.839450e+04	-1.764400e+04
Urban_population	63.0	1.248416e+07	6.286809e+06	3.455253e+06	6.676732e+06	1.227340e+07	1.738873e+07	2.419635e+07
Urban_population_growth	63.0	3.157277e+00	1.018385e+00	1.823103e+00	2.147888e+00	3.173107e+00	4.235305e+00	4.421123e+00
Rural_population_growth	63.0	7.714920e-01	7.745201e-01	-4.652992e-01	8.883368e-03	6.787391e-01	1.422652e+00	1.929147e+00
Rural_population	63.0	1.206130e+07	1.664116e+06	8.314521e+06	1.086419e+07	1.271469e+07	1.348573e+07	1.364485e+07
Population	63.0	2.454545e+07	7.808452e+06	1.176977e+07	1.754092e+07	2.498808e+07	3.103283e+07	3.745797e+07
Population_growth	63.0	1.879790e+00	6.156993e-01	1.023392e+00	1.308207e+00	1.683972e+00	2.558742e+00	2.666415e+00
Population[age>65]	63.0	1.068528e+06	6.981082e+05	3.287510e+05	5.197230e+05	7.894070e+05	1.497293e+06	2.891937e+06
Population[15<age<64]	63.0	1.465653e+07	6.040186e+06	6.086153e+06	9.037042e+06	1.419910e+07	2.013635e+07	2.460810e+07
Mortality_rate_infant	63.0	7.014127e+01	4.196422e+01	1.520000e+01	3.270000e+01	6.130000e+01	1.092500e+02	1.460000e+02

Distributions :



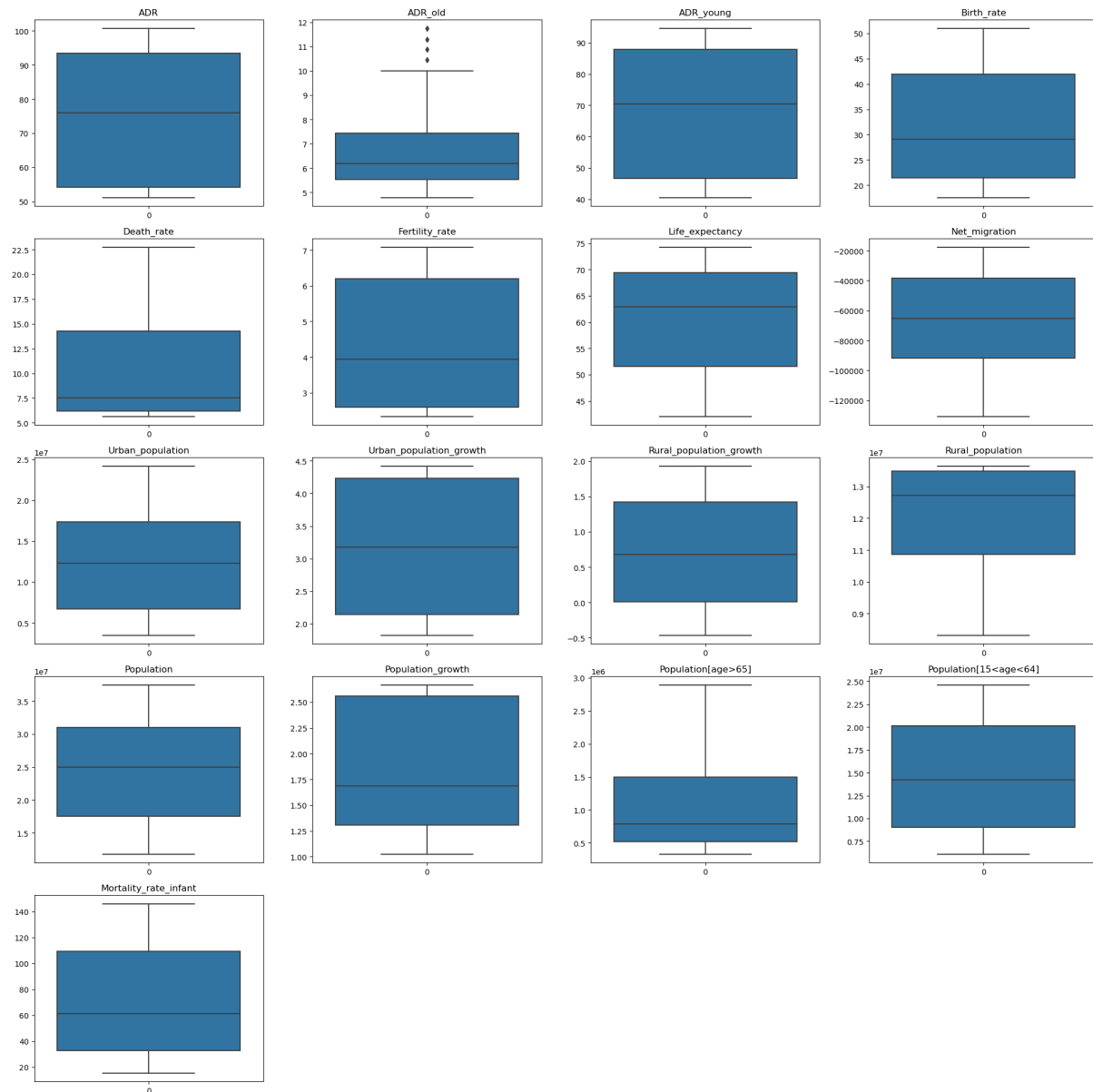
Voir les interprétations dans le notebook

Evolution en fonction du temps :



Voir les interprétations dans le notebook

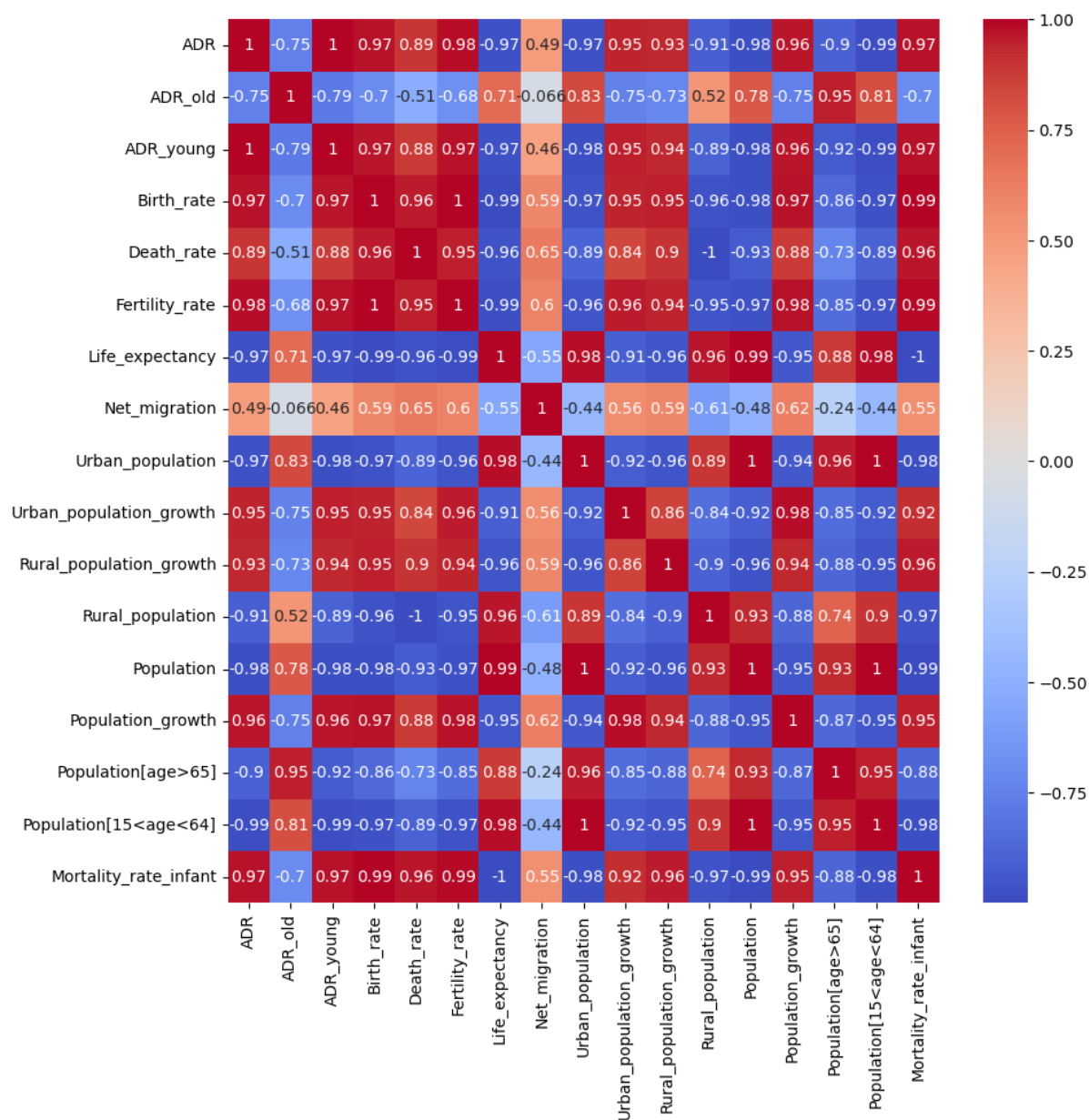
Boxplot :



On remarque que les données sont standardiser , pas de outliers

Analyse bivariée :

Heatmap :

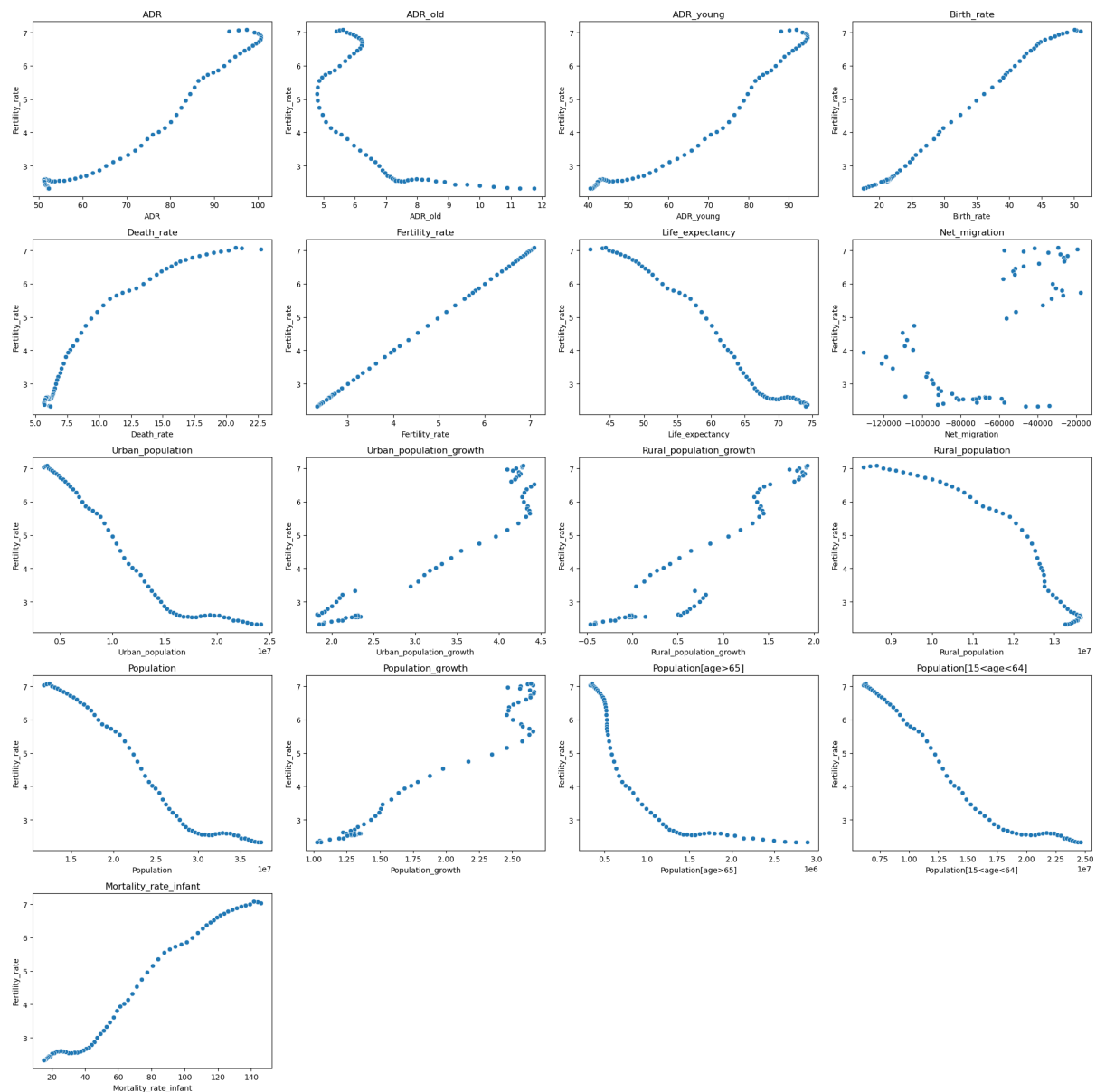


Feature “X” vs Fertility Rate :

feature	correlation
ADR	0.98
ADR_young	0.97
Birth_rate	0.99
Death_rate	0.95
Life_expectancy	0.98
Urban_population	-0.95
Urban_population_growth	0.95
Rural_population_growth	0.94

Rural_population	-0.95
Population	-0.97
Population_growth	0.97
Population[age>65]	-0.84
Population[15<age<64]	-0.96
Mortality_rate_infant	0.98

Scatter plot : Feature “X” vs Fertility Rate



Présélection des variables à fortes corrélation pour la regression :

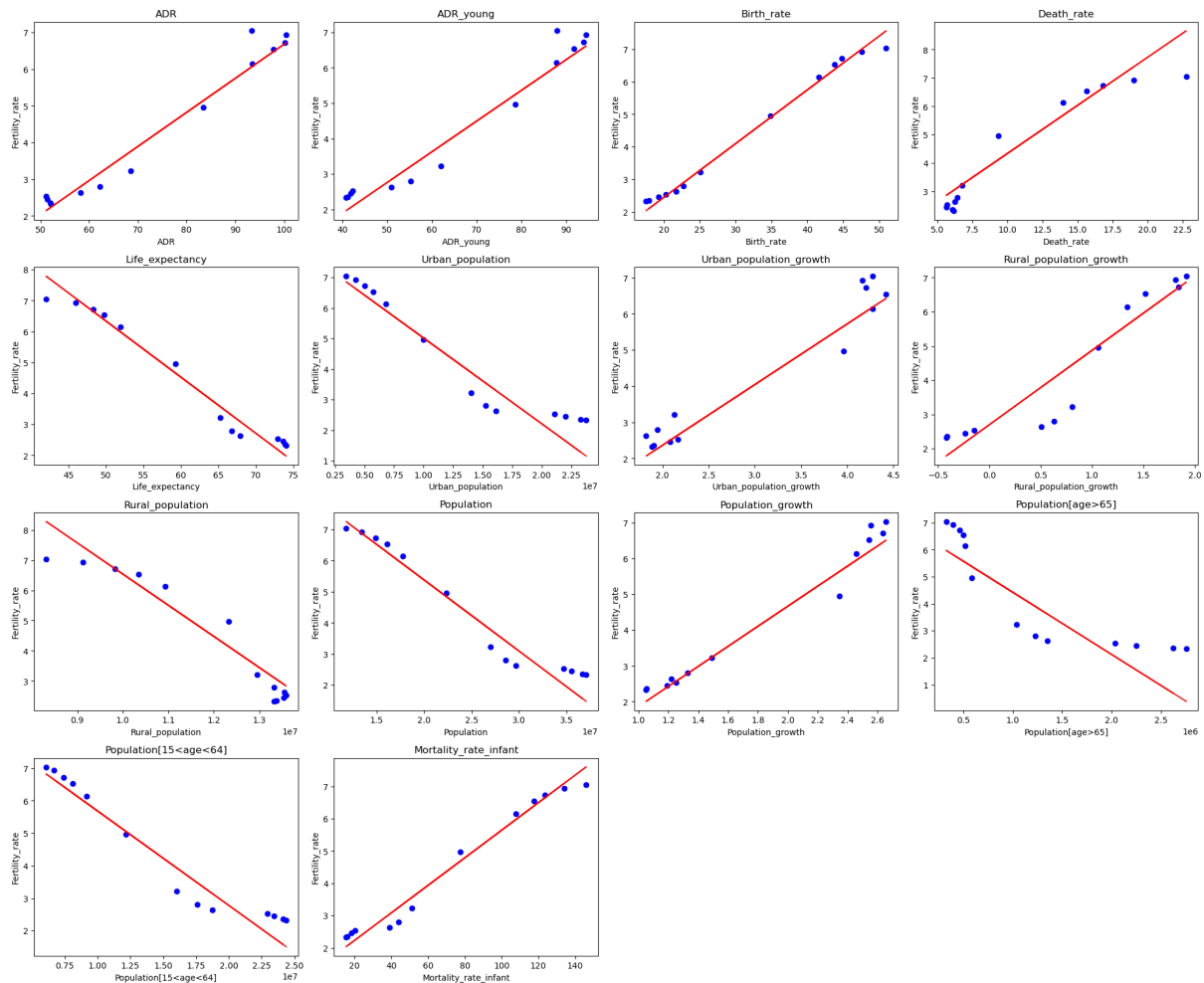
- ADR : 0.98
- ADR_young : 0.97
- Birth_rate : 0.99
- Death_rate : 0.95
- Life_expectancy : 0.98
- Urban_population : -0.95
- Urban_population_growth : 0.95
- Rural_population_growth : 0.94
- Rural_population : -0.95
- Population : -0.97
- Population_growth : 0.97
- Population[age>65] : -0.84
- Population[15<age<64] : -0.96
- Mortality_rate_infant : 0.98

RLS :

RLS avec les attributs sélectionnées :

reg : Variable X vs "Fertility_rate"	Coefficient	Intercept	MSE	R2
ADR	0.09	-2.61	0.14	0.96
ADR_young	0.09	-1.57	0.18	0.95
Birth_rate	0.17	-0.87	0.04	0.99
Death_rate	0.34	0.95	0.41	0.89
Life_expectancy	-0.18	15.40	0.14	0.96
Urban_population	-0.00	7.83	0.42	0.89
Urban_population_growth	1.68	-0.99	0.28	0.93
Rural_population_growth	2.17	2.70	0.45	0.88
Rural_population	-0.00	16.84	0.35	0.91
Population	-0.00	9.9	0.25	0.94
Population_growth	2.81	-0.95	0.13	0.97
Population[age>65]	-0.00	6.73	1.23	0.68
Population[15<age<64]	-0.00	8.59	0.28	0.93
Mortality_rate_infant	0.04	1.38	0.10	0.97

Visualisation :



Sélection manuelle :

1. Taux de natalité (Birth Rate) :

- Il a un coefficient significatif de 0.17 et un R2 très élevé de 0.99, ce qui en fait un prédicteur très fort du taux de fertilité

2. Population Growth :

- Avec un coefficient important de 2.81 et un R2 élevé de 0.97, la croissance de la population est également un prédicteur puissant du taux de fertilité.

3. Age Dependency Ratio (ADR) :

- Bien que son coefficient soit relativement faible à 0.09, l'ADR a un R2 élevé de 0.96, ce qui suggère une forte corrélation avec le taux de fertilité.

4. Life Expectancy :

- L'espérance de vie a un coefficient significatif de -0.18 et un R2 élevé de 0.96, ce qui en fait un facteur important pour expliquer le taux de fertilité.

Ces caractéristiques sont les plus explicatives de l'évolution du taux de fertilité et devraient être prioritaires lors de l'analyse de ce phénomène.

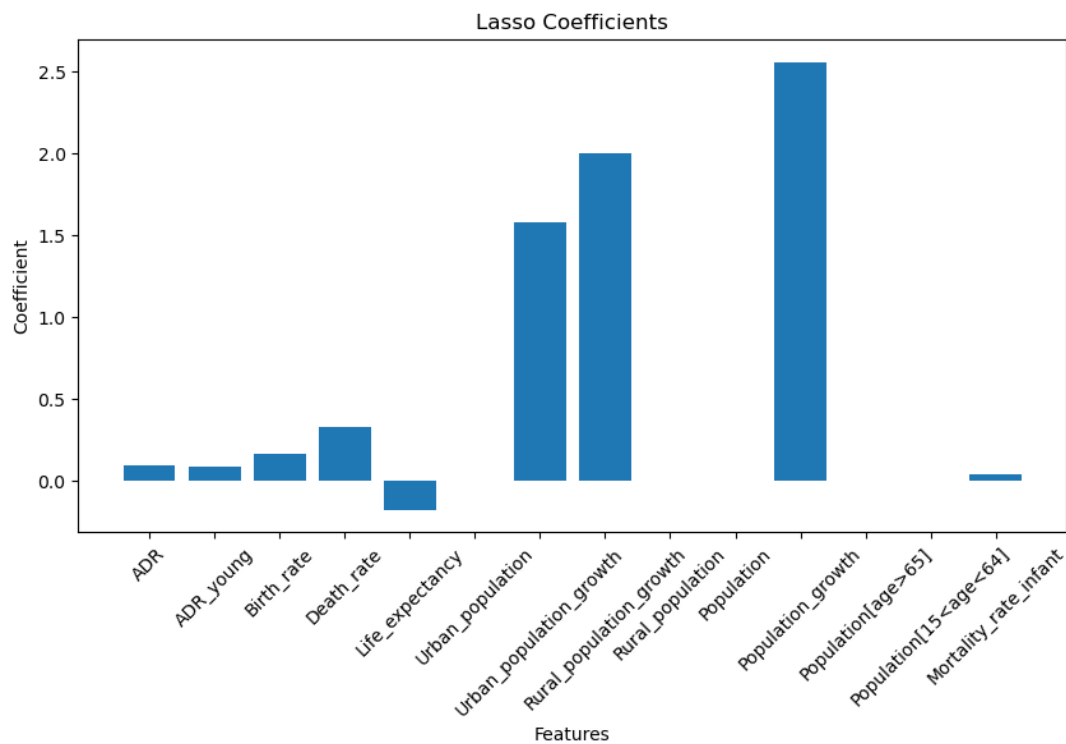
Cross validation :

- ADR: [-12.33191415 0.95520513 -0.54301221 -6.72931766 -31.41077895]
 - Mean: -10.01, Std: 11.70
- ADR_young: [-18.75988855 0.91872642 -1.08185242 -8.36843942 -59.54199557]
 - Mean: -17.37, Std: 22.19
- Birth_rate: [-1.7363277 0.85443654 0.98371777 -2.48724006 -2.87348772]
 - Mean: -1.05, Std: 1.65
- Death_rate: [-131.96576 -3.63266233 -0.02228966 -30.2209869 -41.1357628]
 - Mean: -41.40, Std: 47.89
- Life_expectancy: [-3.12879448 0.09726759 0.71185918 -28.05218377 -28.28509678]
 - Mean: -11.73, Std: 13.48
- Urban_population: [-15.68820871 0.58682739 -0.74862404 -65.56460024 -282.47911213]
 - Mean: -72.78, Std: 107.57
- Urban_population_growth: [-50.61107336 -2.90980441 -0.06961375 -16.95682339 -4.61799378]
 - Mean: -15.03, Std: 18.70
- Rural_population_growth: [-7.61864687 0.16277961 -1.36306152 -110.03644438 -40.02859406]
 - Mean: -31.78, Std: 41.76
- Rural_population: [-135.89634991 -4.22386519 0.47184643 -9.57544889 -41.15130228]
 - Mean: -38.08, Std: 51.03
- Population: [-1.50102832e+00 2.84384974e-01 -9.38084012e-03 -4.28119925e+01 -1.25328989e+02]
 - Mean: -33.87, Std: 48.59
- Population_growth: [-25.56353586 -2.95071324 0.92538254 -1.21816029 -6.22737249]
 - Mean: -7.01, Std: 9.57
- Population[age>65]: [-101.456999 -0.65053448 -5.47524922 -163.275373 -1227.93693]
 - Mean: -299.76, Std: 468.09
- Population[15<age<64]: [-12.85664921 0.66773131 -1.07467971 -36.94802707 -176.82773282]
 - Mean: -45.41, Std: 67.07

- Mortality_rate_infant: [-3.705307 0.14856742 0.75811483 -20.52353311 -18.38287306]
 - Mean: -8.34, Std: 9.23

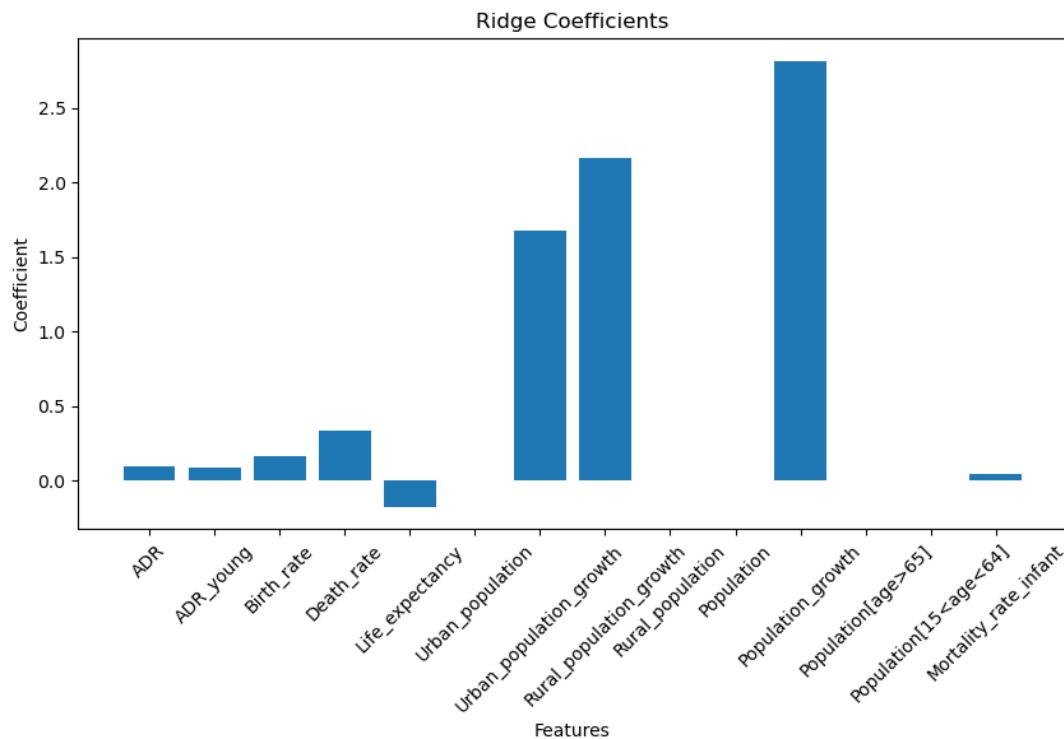
Régularisation :

Lasso :



En conclusion, les caractéristiques les plus importantes pour prédire le taux de fertilité, selon les coefficients Lasso, sont le taux de natalité, la croissance de la population, le taux de mortalité, et l'espérance de vie. Les populations urbaine, rurale, et les tranches d'âge spécifiques ne semblent pas avoir beaucoup d'impact sur le taux de fertilité selon ces résultats.

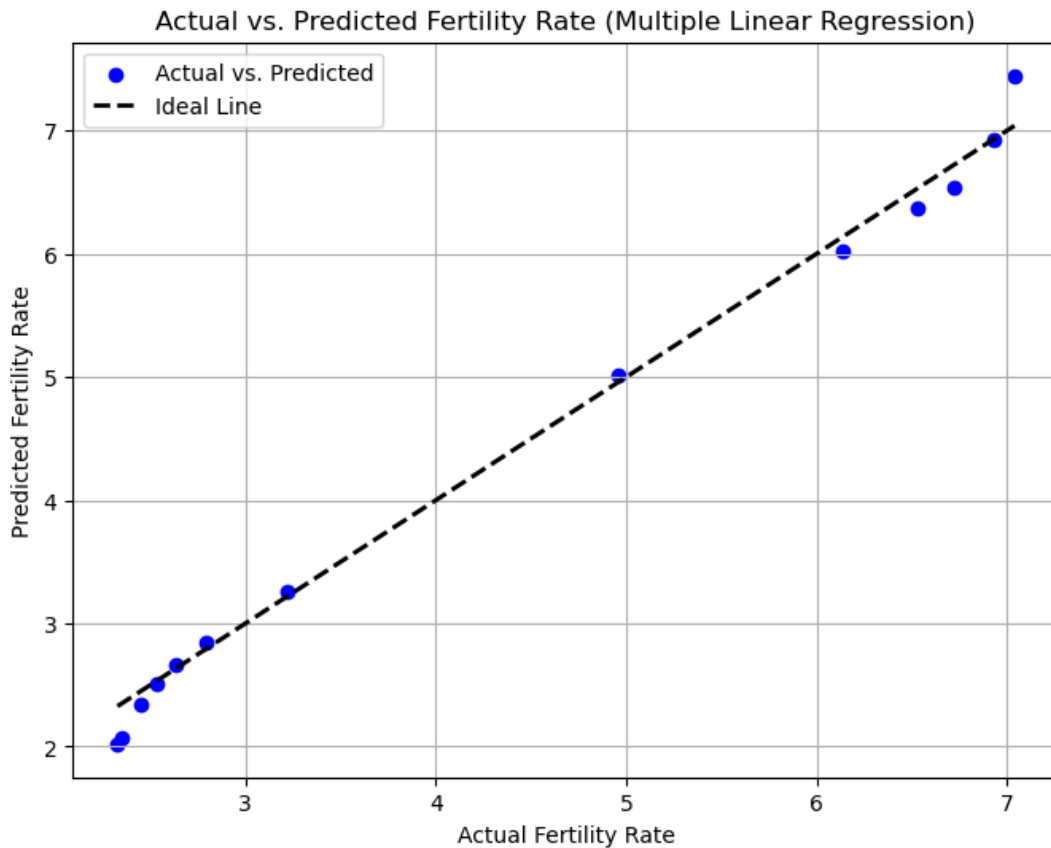
Ridge :



En conclusion, les caractéristiques les plus importantes pour prédire le taux de fertilité, selon les coefficients Ridge, sont le taux de natalité, la croissance de la population, le taux de mortalité, et l'espérance de vie. Les populations urbaine, rurale, et les tranches d'âge spécifiques ne semblent pas avoir beaucoup d'impact sur le taux de fertilité selon ces résultats.

RLM:

- Features :
 - population growth
 - birth rate
 - Death rate
 - life expectancy
- Coefficients: [0.15900948 0.0134182 0.33425347]
- Intercept: -2.1120412155839245
- Mean Squared Error: 0.032888870217987434
- R2 Score: 0.9913597346323106



Les résultats de cette régression linéaire multiple indiquent que les caractéristiques sélectionnées (taux de natalité, espérance de vie et croissance de la population) ont un fort pouvoir explicatif sur le taux de fertilité, comme en témoignent le faible MSE (erreur quadratique moyenne) et le R2 score élevé.

1. Coefficients:

- Pour chaque augmentation d'une unité du taux de natalité, le taux de fertilité augmente d'environ 0.159.
- Pour chaque augmentation d'une unité de l'espérance de vie, le taux de fertilité augmente d'environ 0.0134.
- Pour chaque augmentation d'une unité de la croissance de la population, le taux de fertilité augmente d'environ 0.334.

2. Intercept:

- L'intercept indique la valeur attendue du taux de fertilité lorsque toutes les caractéristiques sont nulles. Dans ce cas, il est d'environ -2.11.

3. Mean Squared Error (MSE):

- Le MSE mesure l'erreur moyenne des prédictions du modèle. Un MSE faible (0.03 dans ce cas) indique que les prédictions du modèle sont proches des valeurs réelles, ce qui suggère une bonne qualité de l'ajustement du modèle aux données.

4. R2 Score:

- Le R2 score (0.99 dans ce cas) représente la proportion de variance dans la variable dépendante (taux de fertilité) qui est expliquée par les variables indépendantes (taux de natalité, espérance de vie, croissance de la population). Un R2 score proche de 1 indique un ajustement très précis du modèle aux données, ce qui est le cas ici.

Extra : Random forest regressor + XGBoost

Random Forest regressor

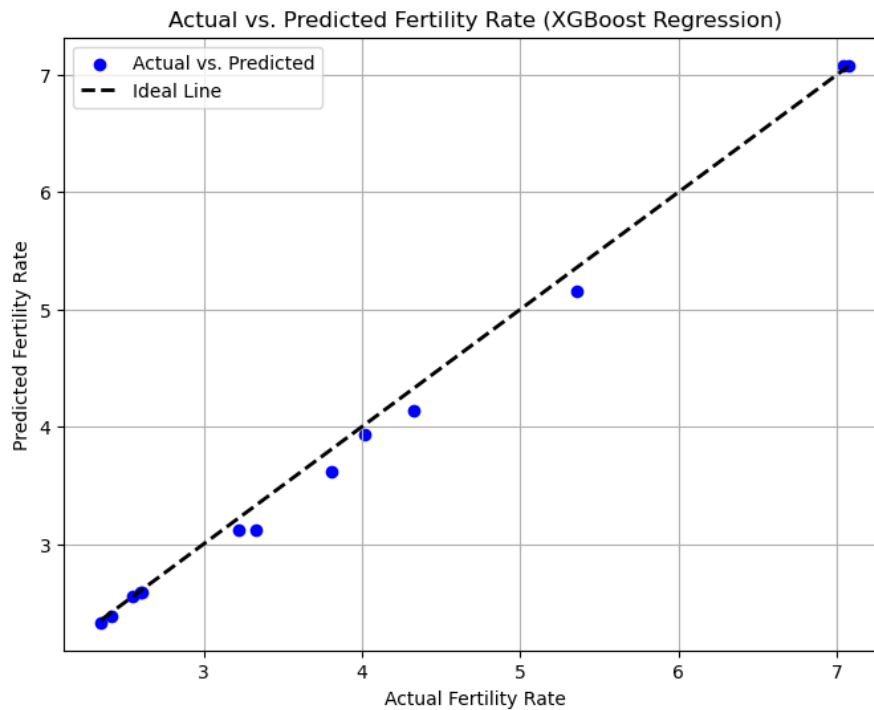
Le "Random Forest Regressor" est un modèle de régression utilisé en apprentissage automatique (machine learning). Il appartient à la famille des méthodes d'ensemble, qui combinent les prédictions de plusieurs modèles individuels pour produire une prédiction finale.

```
MAE: 0.15988366246250463
MSE: 0.03983103885821608
RMSE: 0.19957715014053107
R²: 0.9842203135211276
Cross-validated MAE: 0.17
Cross-validated MSE: -0.17
Cross-validated RMSE: 0.41
Cross-validated R²: -2.15
Cross-validation scores (R-squared): [-0.34145097 -0.25209846 -0.24125283 -0.00555304 -0.01976585]
```

MLS reste plus performant que Random Forest en comparant les metrique et cross validation

XGBoost:

-XGBoost Regression est un algorithme d'apprentissage automatique qui utilise une approche en cascade pour créer un ensemble de modèles prédictifs, principalement des arbres de décision, afin de résoudre des problèmes de régression. Il optimise un objectif de perte spécifique, tel que la perte quadratique, en ajustant itérativement les modèles pour minimiser les erreurs de prédiction



MAE: 0.08486364203232985

MSE: 0.013483327631425976

RMSE: 0.11611773177007022

R^2 : 0.9946583697333836

Resultat de Cross validation:

Cross-Validation Mean MSE: 0.026314445044790274

Cross-Validation Standard Deviation of MSE: 0.014757361400134884

En résumé, ces résultats suggèrent que le modèle XGBoost avec les caractéristiques sélectionnées (taux de natalité, espérance de vie, croissance de la population) est capable de bien expliquer la variance du taux de fertilité, avec un MSE faible et un R^2 score élevé. Cela indique que ces caractéristiques sont des prédicteurs significatifs du taux de fertilité et qu'elles fournissent des informations précieuses pour comprendre et prédire ce phénomène.

Conclusion :

Merci Madame El Ghali pour cette opportunité qui nous a permis de mettre en pratique les concepts appris dans notre cours Modèles de régression dans le cadre du module : “**Modèles de regression et série chronologiques :**”. Nous avons ainsi acquis des compétences utiles pour la conception et la

mise en place d'un workflow cohérent et complet de ces méthodes importantes des sciences des données.