

Projet Info S7

Création d'une plateforme de traitement
de données statistiques



Lou Pelissolo

Département Informatique



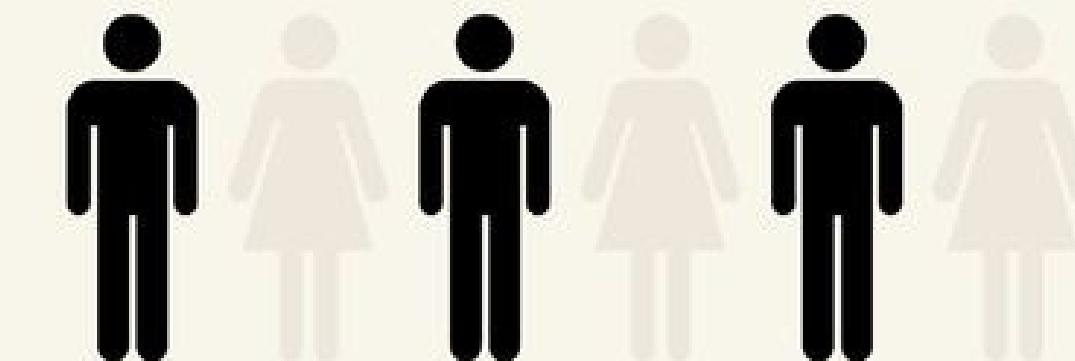
Tutrice : Dominique Benmouffek

CAROLINE CRIADO PEREZ



FEMMES INVISIBLES

COMMENT LE MANQUE DE DONNÉES
SUR LES FEMMES DESSINE UN MONDE
FAIT POUR LES HOMMES



POCKET

Plan de la présentation *

1

Présentation des données

- Le "Gender Development Index" (GDI)
- Le "Gender Inequality Index"(GII)

2

La plateforme Streamlit

- L'outil Streamlit
- Les différents onglets: informations, visualisation, corrélation, prédictions...

3

Analyse des résultats

- Analyse des prédictions
- Fiabilité des résultats
- Comparaison avec le HDI

1

Choix des jeux de données



Sources : Kaggle, Data gouv, UNDP

GDI

Gender Development Index

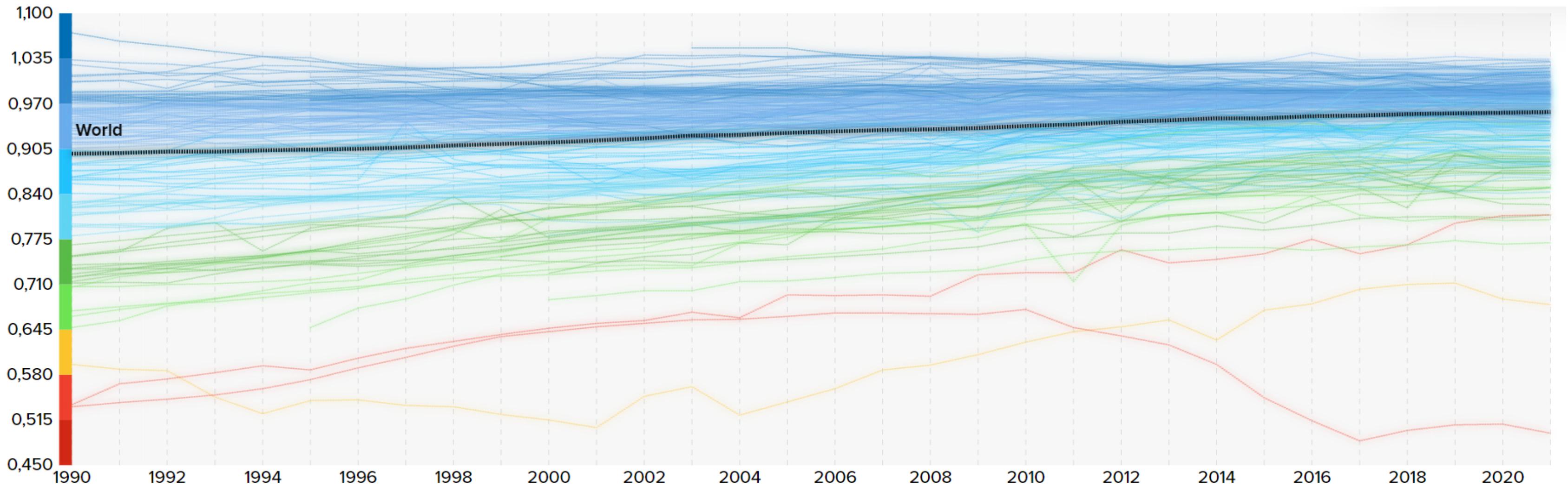
GII

Gender Inequality Index

HDI

Human Development Index

Le Gender Development Index (GDI)



Calcul

< 1,04

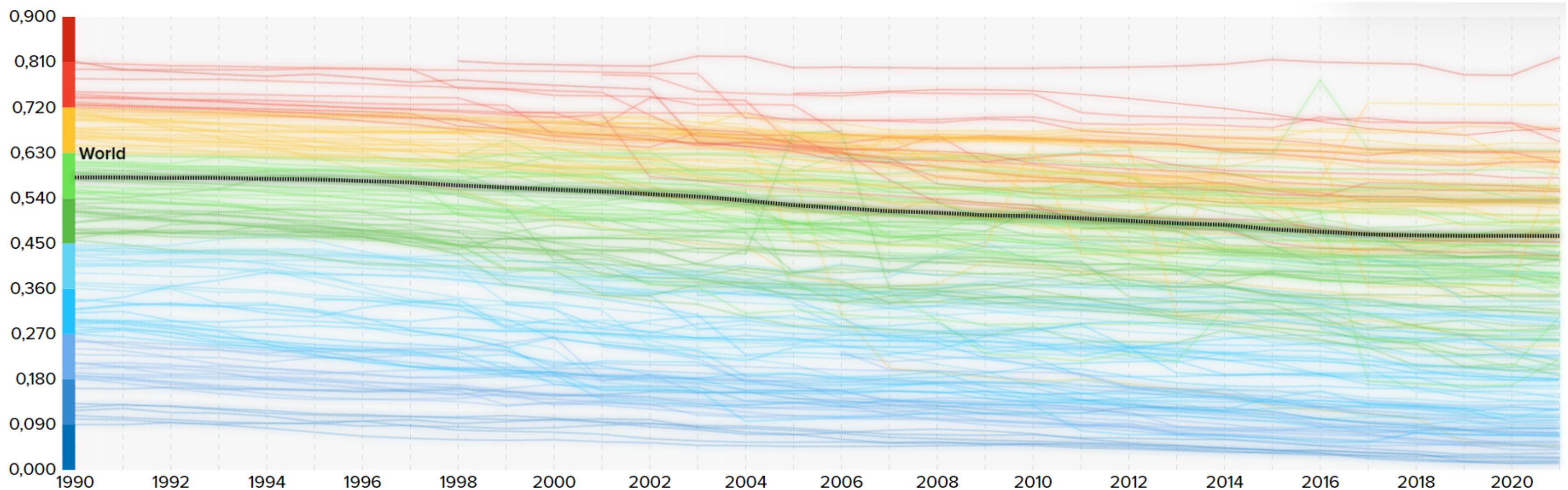
Santé: espérance de vie à la naissance

Education : moyennes d'années de scolarisation

Contrôle sur les ressources économiques: RNB (Revenu National Brut)



Le Gender Inequality Index (GII)



Calcul

<1

Santé "reproductive" : mortalité maternelle, taux de natalité chez les adolescentes

Autonomie: pourcentage de femmes au parlement, dans l'éducation secondaire

Marché du travail : participation des femmes à la force de travail

2

La plateforme Streamlit

Streamlit : un module de Python qui permet la création d'applications graphiques sous la forme d'interfaces Web

Les différents onglets

Menu

Sélectionnez l'onglet :

- Informations et Résumé
- Corrélations
- Visualisation
- Classification
- Prédictions pour le GDI

Informations et résumé

- Affichage des 70 premières lignes de jeu de données sélectionné
- Statistiques sur les variables quantitatives (moyenne, minimum, maximum...)

Statistiques sur les variables quantitatives

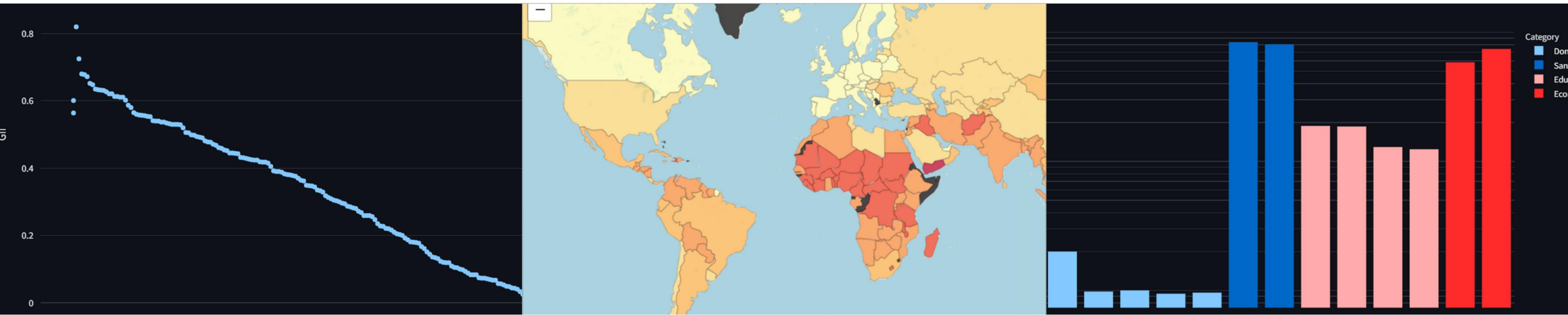
	Variable	Min	Max	Moyenne
0	GII	0.013	0.82	0.3444
1	Rank	1	170	85.3765
2	Maternal_mortality	2	1,150	160.0272
3	Adolescent_birth_rate	1.6	170.5	44.5979
4	Seats_parliament	0	55.7	24.7016
5	F_secondary_educ	6.4	100	62.7068
6	M_secondary_educ	13	100	67.0684
7	F_Labour_force	6	83.1	50.2244
8	M_Labour_force	43.9	95.5	69.8633



Corrélations

- Matrices de corrélation
- Calcul des p-values et analyse de corrélation

Visualisation



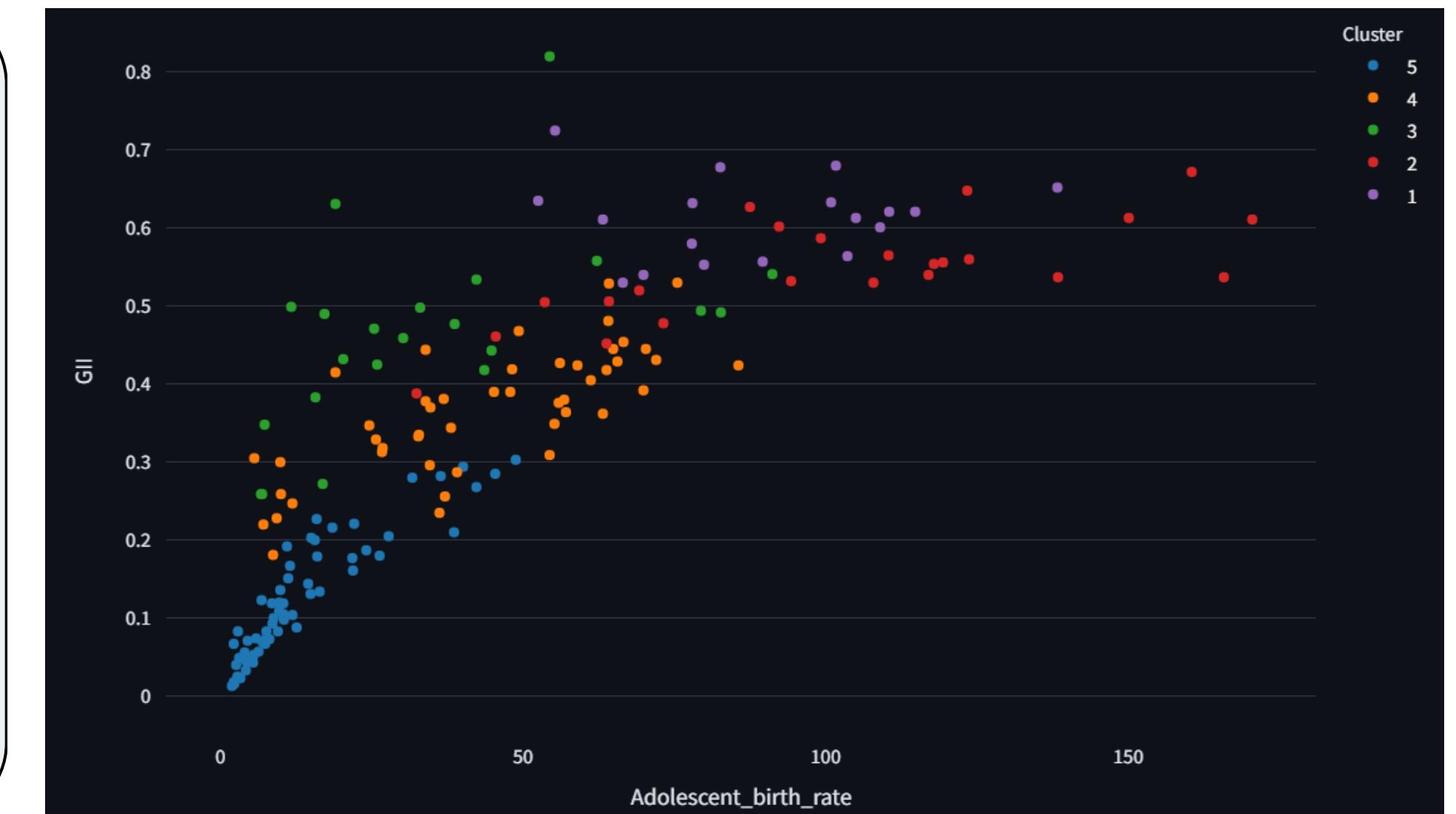
Nuage de
points

Cartes du
monde

Histogramme

Classification

- Création de groupes de pays par l'algorithme des **K-moyennes**
- Affichage de la répartition des clusters en fonction des variables choisies
- Affichage des profils de pays sur une carte



Prédictions

Le modèle ARIMA



Un modèle **ARIMA**(p,q,d) est particulièrement adapté à l'analyse des séries temporelles en raison de sa capacité à capturer les tendances, les saisons et les variations autoregressives.



Auto-Regressive (AR):

On suppose que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents (+ un terme d'erreur)
p : nombre de termes auto-regressifs



Integrated (I):

On suppose que chaque point présente une différence constante avec le point précédent
d : nombre de différences



Moving Average (MA):

On suppose que chaque point est fonction des erreurs entachant les points précédents (+ sa propre erreur)
q : nombre de moyennes mobiles

③ Analyse des résultats



Critique des indicateurs et comparaison



Analyse des profils de pays



Comparaison avec l'indice de développement humain (HDI)

Problèmes du GDI

Prise en compte de l'**espérance de vie** :

- les femmes ont naturellement une espérance de vie plus élevée que les hommes
- meurtre != féminicide



Prise en compte du "**GNI**":

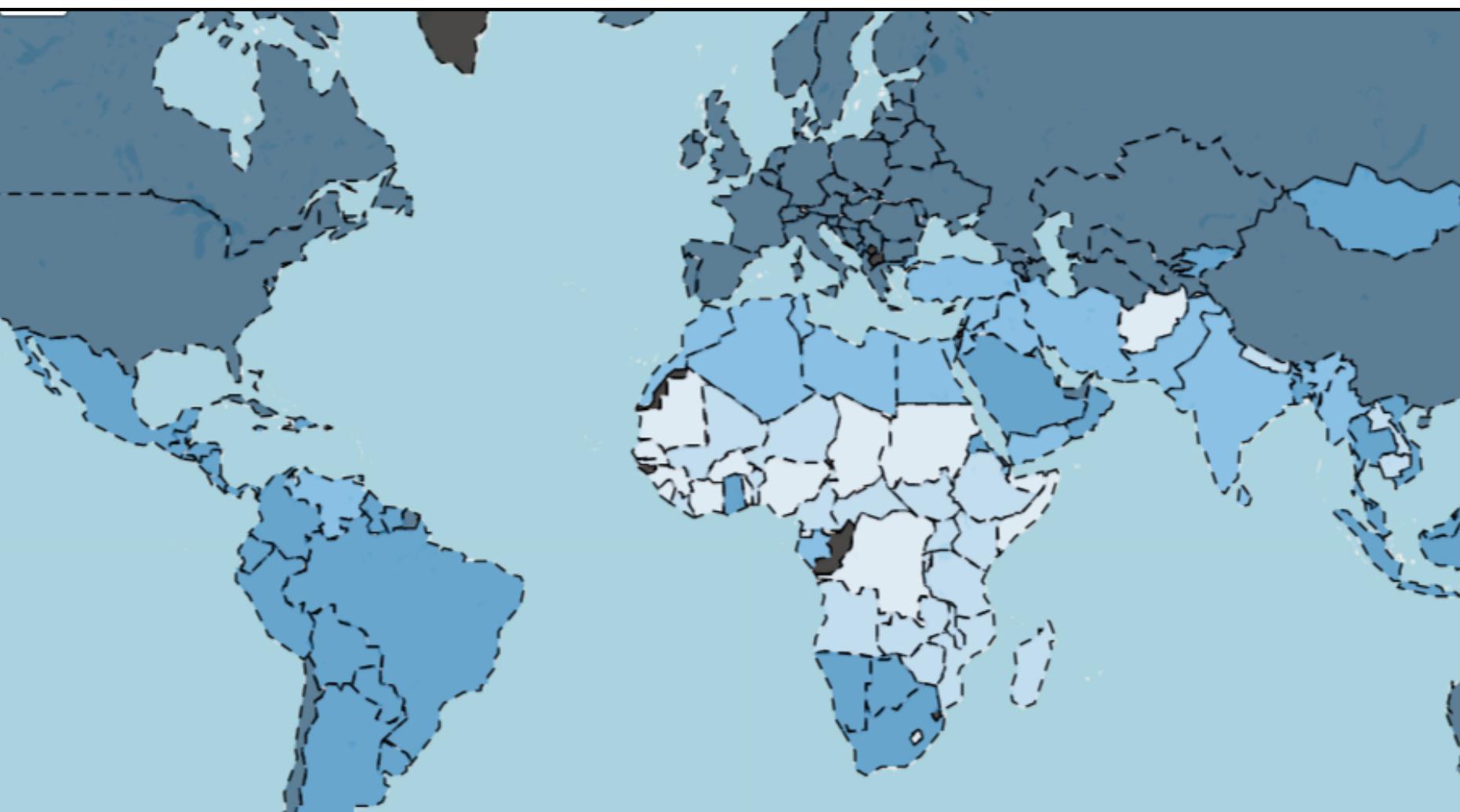
- déséquilibre entre les pays riches et les pays pauvres

Le GII semble globalement être un meilleur indicateur



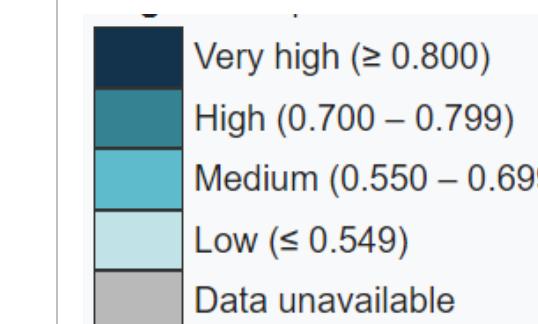
L'aspect économique du GII calcule le % de femmes participant à la force de travail, ce qui ne rend pas compte des salaires plus faibles, ou du taux plus élevé de travail domestique non rémunéré

Analyse des profils de pays



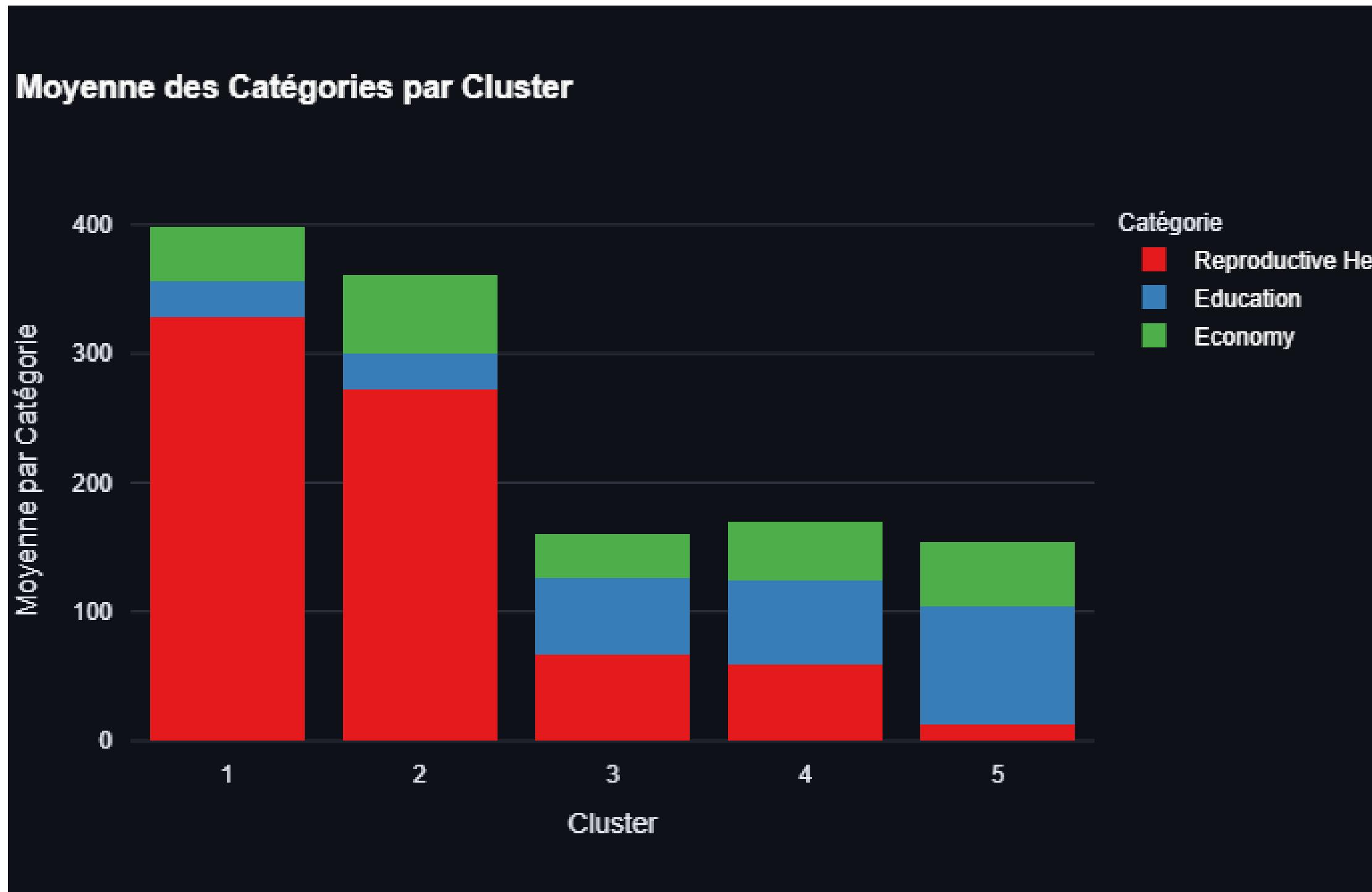
Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

Profils de pays formés par l'algorithme K-means



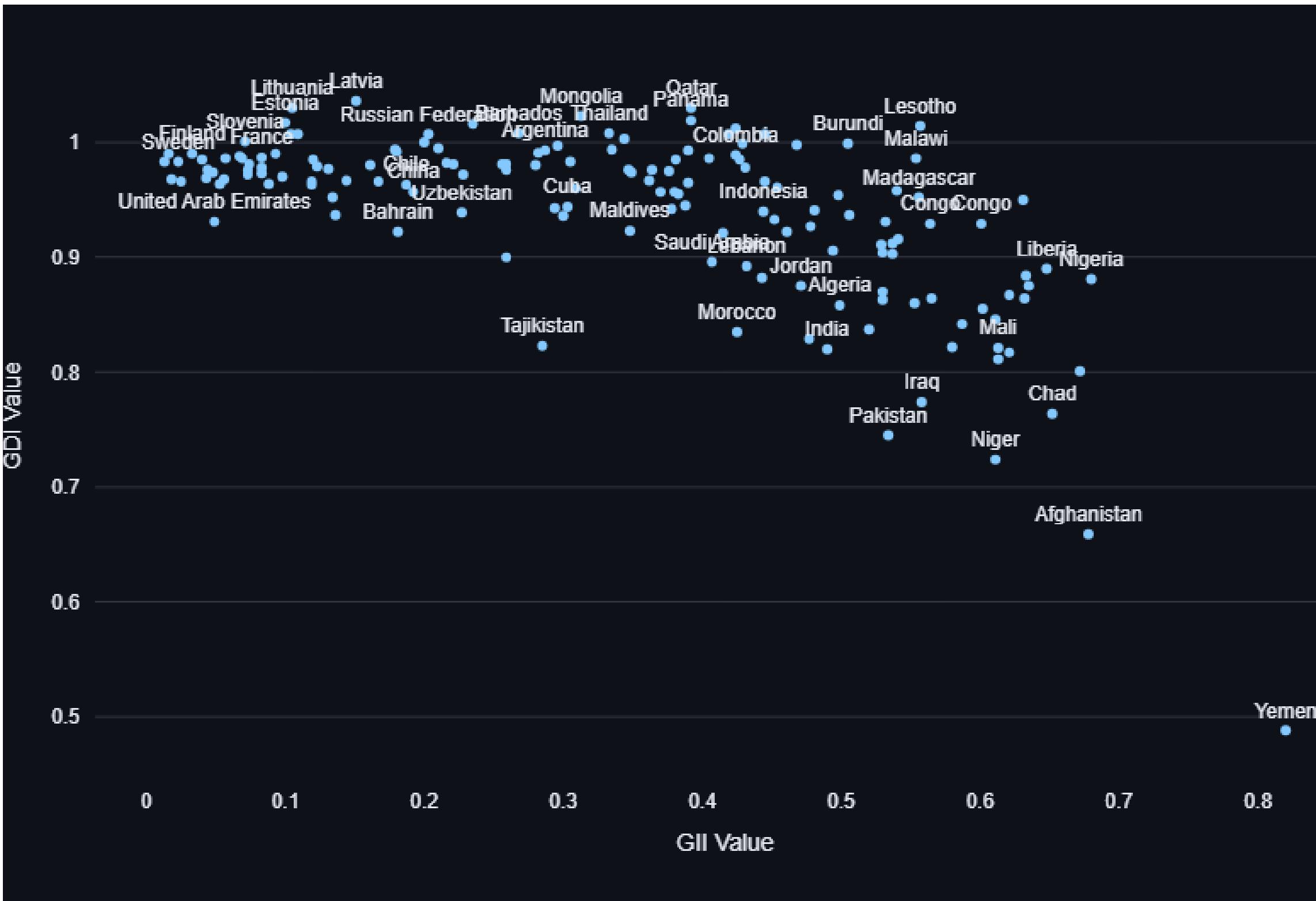
Carte de l'ONU
typologie des pays en fonction de leur Indice de
Développement Humain (IDH)

L'importance du facteur "Santé Reproductive"



- Ecart important entre les pays du cluster 2 et du cluster 3: IDH entre "moyen" et "bas" selon l'ONU
- Pertinence des 5 clusters ?
- Le facteur le plus discriminant: le facteur "Santé reproductive"

Comparaison entre le GDI et le GII



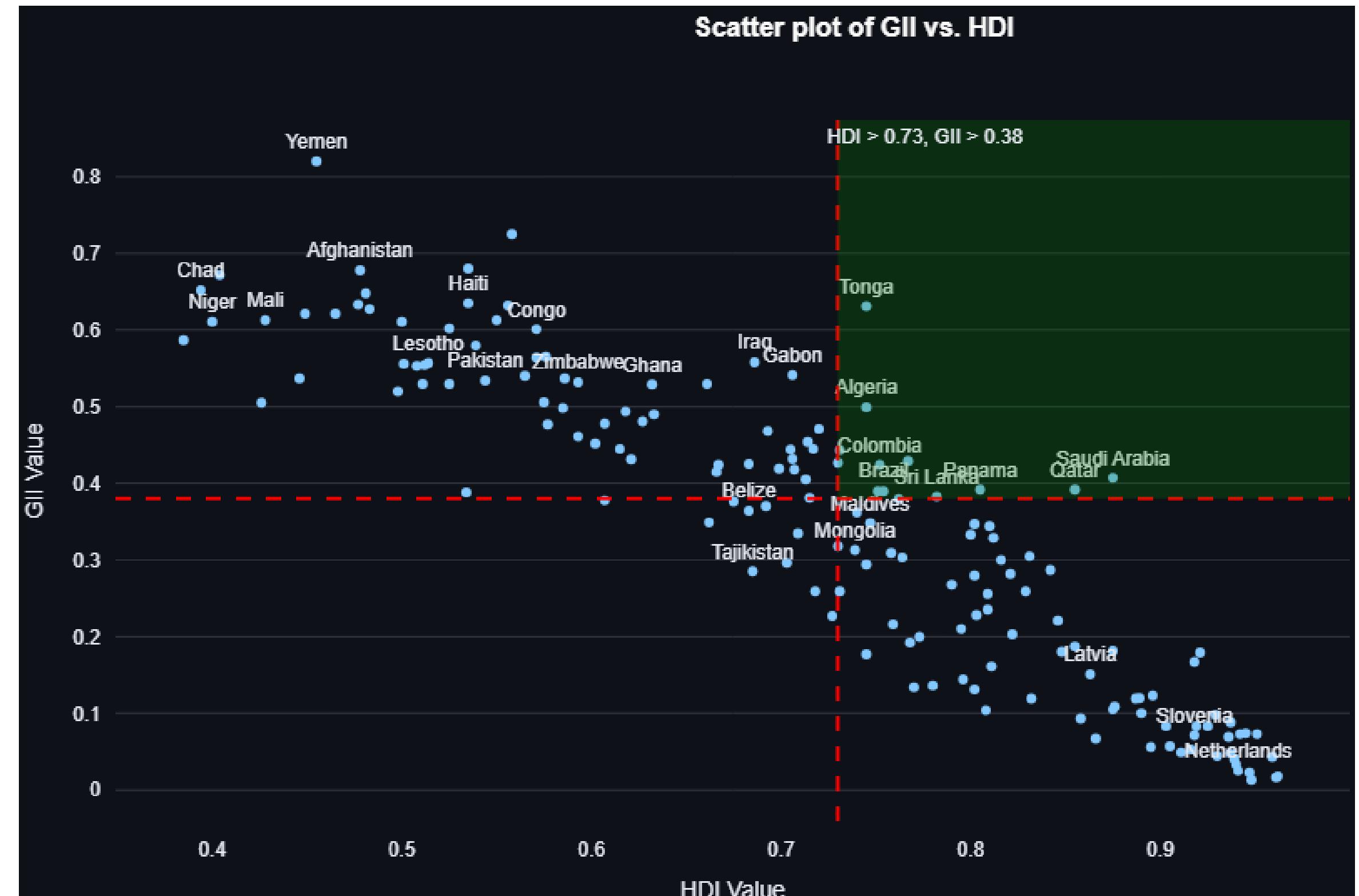
- Des pays comme la Chine, la Moldavie, le Vietnam gagnent jusqu'à 40 places sur le GII par rapport au GDI
-> taux de mortalité maternelle et fécondité chez les adolescentes faibles + niveau élevé de participation des femmes au marché du travail
- L'Arabie Saoudite et le Qatar perdent presque 50 places
-> faible taux de travail chez les femmes, et peu de femmes au parlement

Illustration du problème du GDI sur la prise en compte du RNB

Comparaison entre le GII et le HDI

Globalement, tous les pays ayant l'HDI le plus élevé sont catégorisés comme les moins inégalitaires **mais il y a des exceptions**

Pays avec HDI>0,73 et GII > 0,38:
Algérie, Brésil, Qatar, Arabie Saoudite, Panama, Tonga, Sri Lanka...



Conclusion:

Limites du modèle

- Prédictions actuelles peu concluantes
- La plupart des fonctionnalités ne marchent qu'avec mes deux jeux de données

Problème de calcul des indicateurs actuels

- Les indicateurs les plus utilisées pour calculer les inégalités de genre ne prennent pas en compte **des facteurs très importants** tels que: le travail domestique, la propriété, le taux de violences sexuelles, la participation à la communauté...

Perspectives futures

- Amélioration des prédictions: travail à faire sur la partie **apprentissage automatique**
- Inclure du HTML/CSS pour rendre la plateforme plus esthétique
- Rapport et dépôt GIT