

A serene landscape photograph featuring a calm lake in the foreground, reflecting the surrounding environment. A dense forest of evergreen trees lines the middle ground, and in the background, a range of rugged mountains with patches of snow is visible under a soft, hazy sky. The overall mood is peaceful and majestic.

Happiness Project

Thibaud



INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

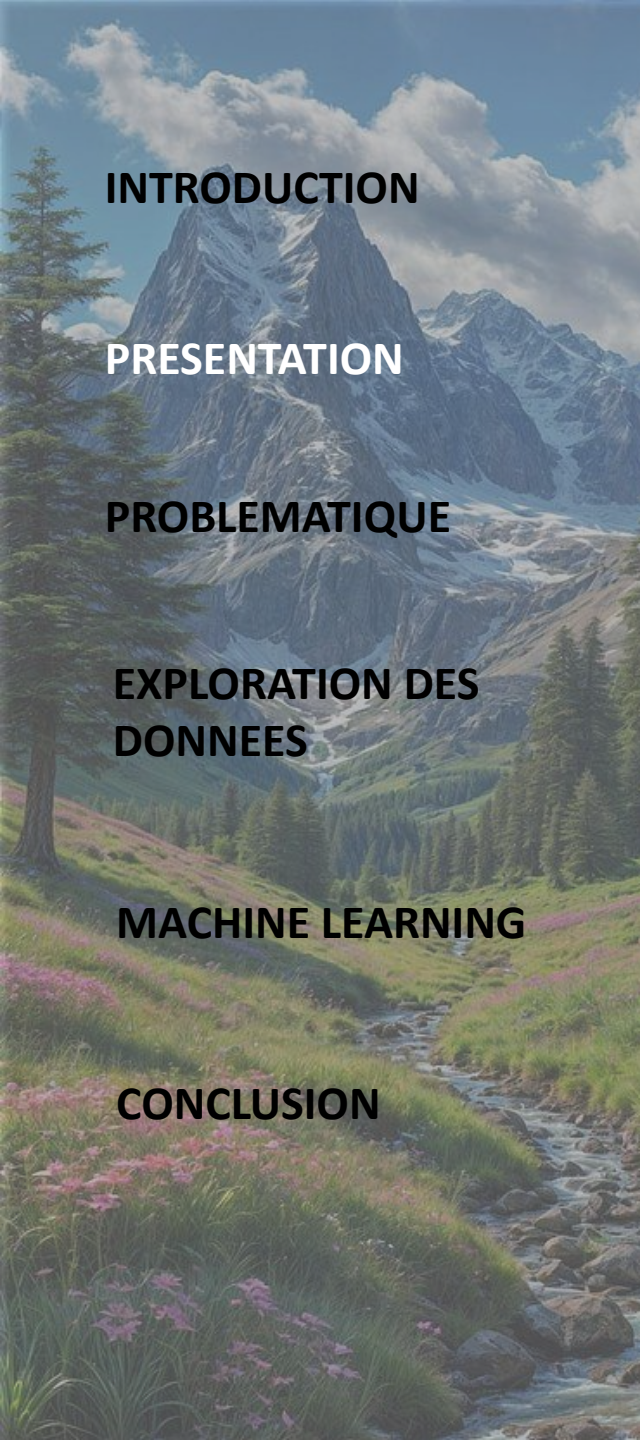
The World Happiness Report (Rapport Mondial sur le Bonheur 2024) est un rapport publié chaque année par Gallup, les Nations unies et l'Université d'Oxford. Il mesure le bonheur des populations à travers le monde et classe les pays en fonction du bien-être de leurs citoyens.

Ce rapport évalue le niveau de bonheur selon un score sur 10. Il fournit également des données normalisées sur un certain nombre d'autres indicateurs socio-économiques.

Nous avons sélectionné le rapport de 2024 ainsi qu'un dataset compilant les données de 2005 à 2023, nous permettant de produire une analyse sur le temps long.

Nous avons utilisé ces données pour analyser **les facteurs qui influencent le bonheur** dans différents pays et comment ces facteurs évoluent d'une année à l'autre.





INTRODUCTION



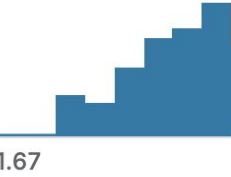
PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

Country name	Regional indicator	Ladder score	upperwhisker	lowerwhisker
Name of the country.	Region the country belongs to.			
143 unique values	Sub-Saharan Afri... 24%			
	Western Europe 14%			
	Other (88) 62%			
Finland	Western Europe	7.741	7.815	7.667
Denmark	Western Europe	7.583	7.665	7.500
Iceland	Western Europe	7.525	7.618	7.433
Sweden	Western Europe	7.344	7.422	7.267
Israel	Middle East and North Africa	7.341	7.405	7.277
Netherlands	Western Europe	7.319	7.383	7.256
Norway	Western Europe	7.302	7.389	7.215
Luxembourg	Western Europe	7.122	7.213	7.031
Switzerland	Western Europe	7.060	7.147	6.973

2 fichiers
23 colonnes au total

[Source : Kaggle](#)

Ladder score : c'est ce qu'on appelle l'échelle de Cantril. Elle demande aux répondants de penser à une échelle, la meilleure vie possible pour eux étant un 10 et la pire vie possible étant un 0. Il leur est ensuite demandé d'évaluer leur propre vie actuelle sur cette note de 0 à 10. Le reste sont des données normalisées dans des formats numériques.



INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

« Comment le niveau de bonheur varie-t-il selon les pays et comment les différents facteurs étudiés l'influencent-ils? »

INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

1. Changement de l'en-tête

	A ^B _C Column1	A ^B _C Column2	A ^B _C Column3	A ^B _C Column4	A ^B _C Column5	A ^B _C Column6
1	Country name	Regional indicator	Ladder score	upperwhisker	lowerwhisker	Ladder score



A ^B _C Country name	A ^B _C Regional indicator	A ^B _C Ladder score	A ^B _C upperwhisker	A ^B _C lowerwhisker
Finland	Western Europe	7.741	7.815	7.667



PowerBI

INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

2. Remplacement des "," en "."



7.741	7.815	7.667
7.583	7.665	7.500
7.525	7.618	7.433
7.344	7.422	7.267
7.341	7.405	7.277
7.319	7.383	7.256
7.302	7.389	7.215
7.122	7.213	7.031
7.060	7.147	6.973
7.057	7.141	6.973
7.029	7.105	6.954
6.955	7.051	6.860
6.951	7.060	6.843
6.905	6.986	6.824
6.900	6.984	6.815
6.894	6.961	6.827
6.838	6.927	6.749
6.822	6.903	6.741
6.818	6.896	6.739
6.749	6.833	6.665
6.743	6.843	6.643
6.733	6.823	6.643



Replace Values

Replace one value with another in the selected columns.

Value To Find

Replace With

> Advanced options

OK Cancel



7.741	7.815	7.667
7.583	7.665	7.5
7.525	7.618	7.433
7.344	7.422	7.267
7.341	7.405	7.277
7.319	7.383	7.256
7.302	7.389	7.215
7.122	7.213	7.031
7.06	7.147	6.973
7.057	7.141	6.973
7.029	7.105	6.954
6.955	7.051	6.86
6.951	7.06	6.843
6.905	6.986	6.824
6.9	6.984	6.815
6.894	6.961	6.827
6.838	6.927	6.749
6.822	6.903	6.741
6.818	6.896	6.739
6.749	6.833	6.665
6.743	6.843	6.643
6.733	6.823	6.643

INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

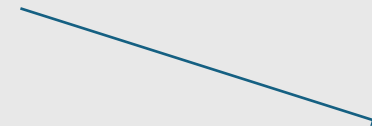
MACHINE LEARNING

CONCLUSION

3. Remplacement des valeurs manquantes en "null"



7.325
9.122



7,325
null
null
9,122

INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

4. Jointures des deux datasets



Merge

Select a table and matching columns to create a merged table.

World-happiness-report-2005-2023

Country name	year	Life Ladder	Log GDP per capita	Social support	Healthy life expectancy at birth	Fre
Afghanistan	2008	3,724		7,35	0,451	50,5
Afghanistan	2009	4,402		7,509	0,552	50,8
Afghanistan	2010	4,758		7,614	0,539	51,1
Afghanistan	2011	3,832		7,581	0,521	51,4

World-happiness-report-2024

Country name	Regional indicator	Ladder score	upperwhisker	lowerwhisker	Log GDP per capita
Finland	Western Europe	7,741	7,815	7,667	1,844
Denmark	Western Europe	7,583	7,665	7,5	1,908
Iceland	Western Europe	7,525	7,618	7,433	1,881
Sweden	Western Europe	7,344	7,422	7,267	1,878

Join Kind

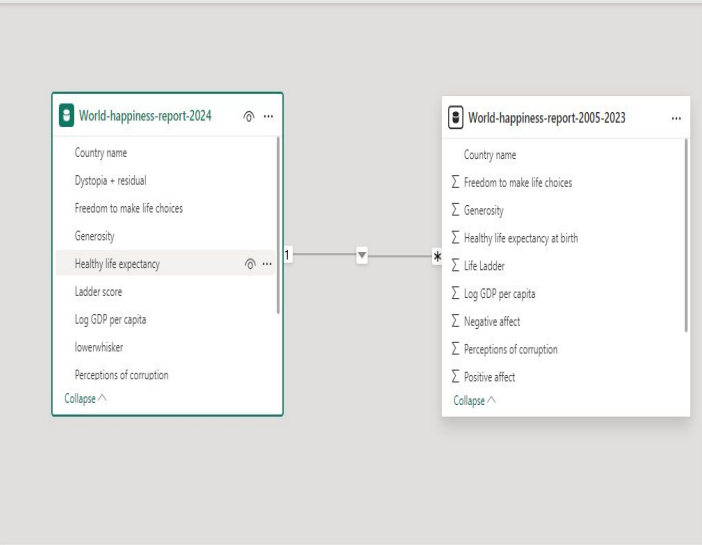
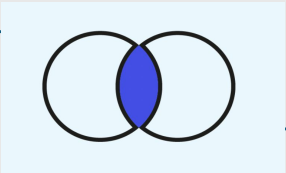
Left Outer (all from first, matching from second)

☐ Use fuzzy matching to perform the merge

Fuzzy matching options

☒ The selection matches 2237 of 2363 rows from the first table.

	1.2 Perceptions of corruption	1.2 Positive affect	1.2 Negative affect		World-happiness-report-2024
0,164	0,882	0,414	0,258	Table	
0,187	0,85	0,481	0,237	Table	
0,118	0,707	0,517	0,275	Table	
0,16	0,731	0,48	0,267	Table	
0,234	0,776	0,614	0,268	Table	
0,059	0,823	0,547	0,273	Table	
0,102	0,871	0,492	0,375	Table	
0,078	0,881	0,491	0,339	Table	
0,04	0,793	0,501	0,348	Table	
-0,123	0,954	0,435	0,371	Table	
-0,095	0,928	0,385	0,405	Table	
-0,109	0,924	0,324	0,502	Table	
-0,085	0,946	0,179	0,607	Table	
null	0,733	0,206	0,576	Table	
null	0,738	0,261	0,46	Table	
-0,013	0,875	0,489	0,246	Table	
-0,162	0,864	0,564	0,279	Table	
-0,176	0,726	0,576	0,3	Table	
-0,209	0,877	0,566	0,257	Table	
-0,173	0,848	0,553	0,271	Table	
-0,131	0,863	0,541	0,338	Table	
-0,029	0,883	0,573	0,335	Table	
-0,085	0,885	0,579	0,35	Table	



INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

5. Ajout d'une colonne "year" dans le dataset 2024



PowerBI

Custom Column

Add a column that is computed from the other columns.

New column name

Custom column formula ⓘ

Available columns

- Country name
- Regional indicator
- Ladder score
- upperwhisker
- lowerwhisker
- Log GDP per capita
- Social support

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK Cancel

General overview for 2024

Countries
143

Average Ladder
score
5.53

Least happy country
Afghanistan

1,72

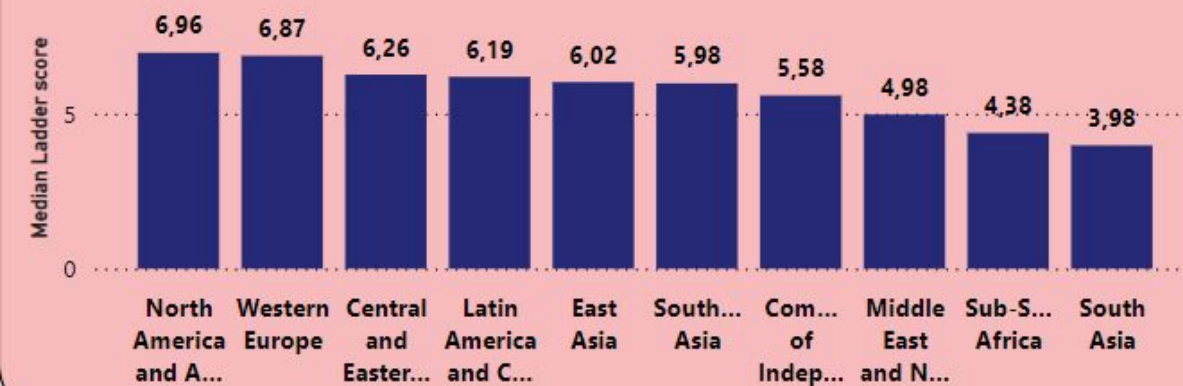
Min Ladder score

Happiest country
Finland

7,74

Max Ladder score

Median happiness score per major region



Regional indicator

Central and Eastern Europe

Commonwealth of Independent States

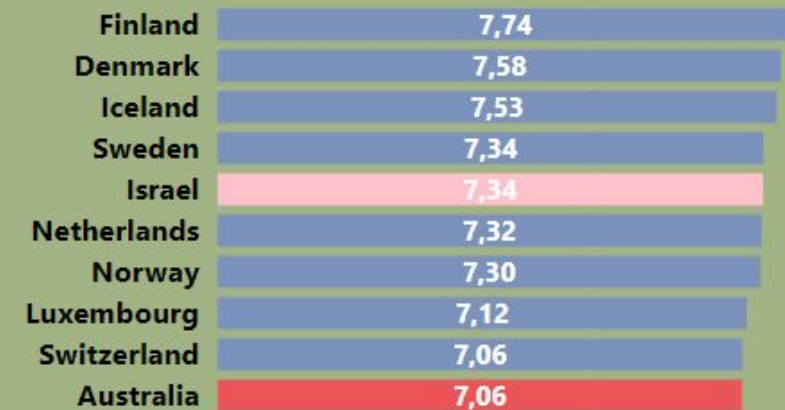
East Asia

Latin America and Caribbean

Ladder Score by country

Regional indicator

- Central and Eastern Europe
- Commonwealth of Independent States
- East Asia
- Latin America and Caribbean
- Middle East and North Africa
- North America and ANZ
- South Asia
- Southeast Asia



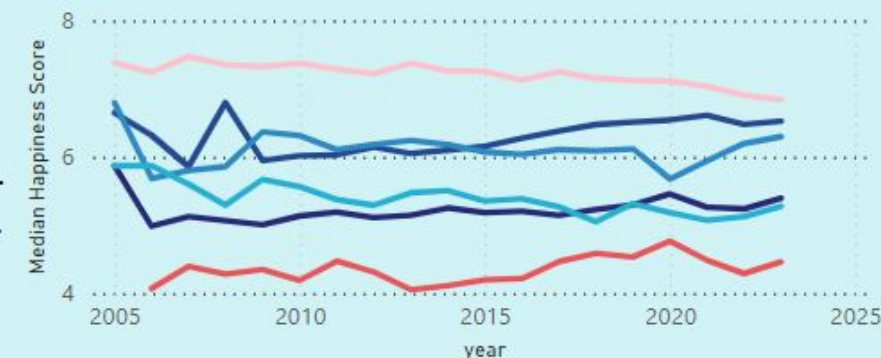
Happiness score accross the Globe



Median Happiness Score from 2005 to 2023

Grouped Regional Indicator

- Asia
- Europe
- Latin America and Caribbean
- Middle East and North Africa
- North America and ANZ
- Sub-Saharan Africa



The evolution of the ladder score across the globe between 2013 and 2023

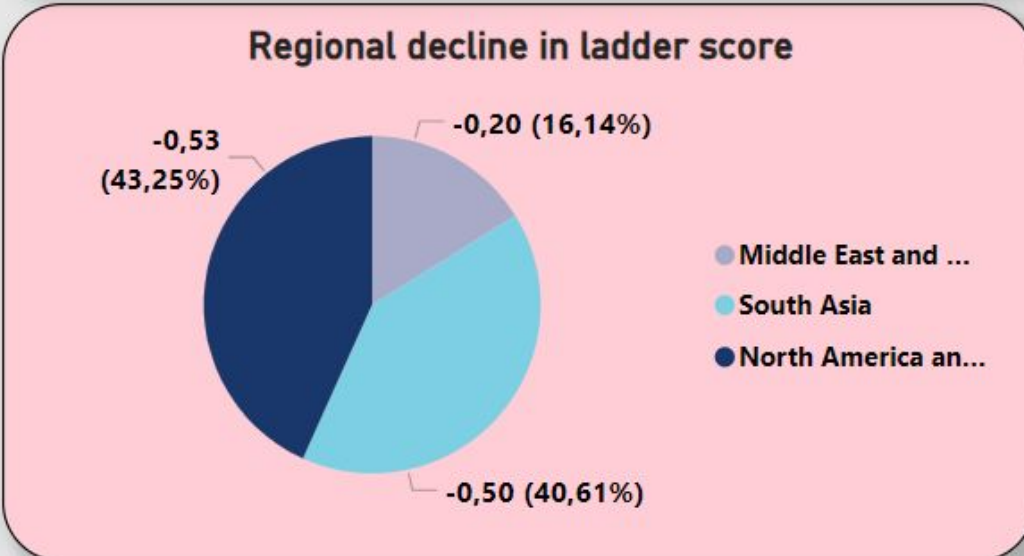
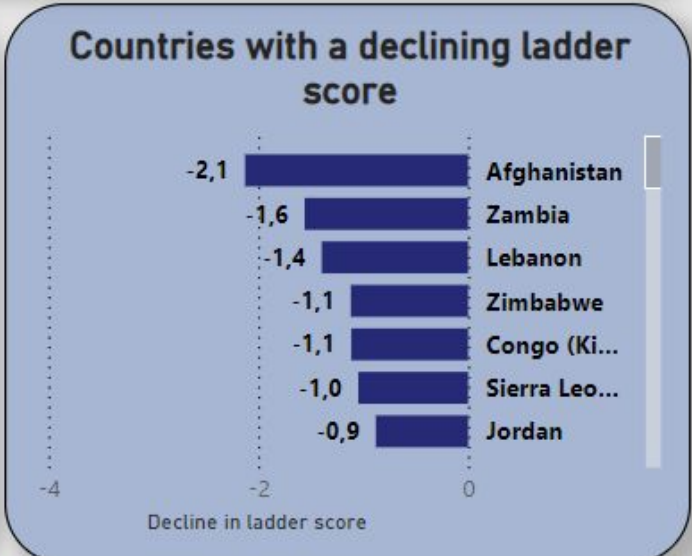
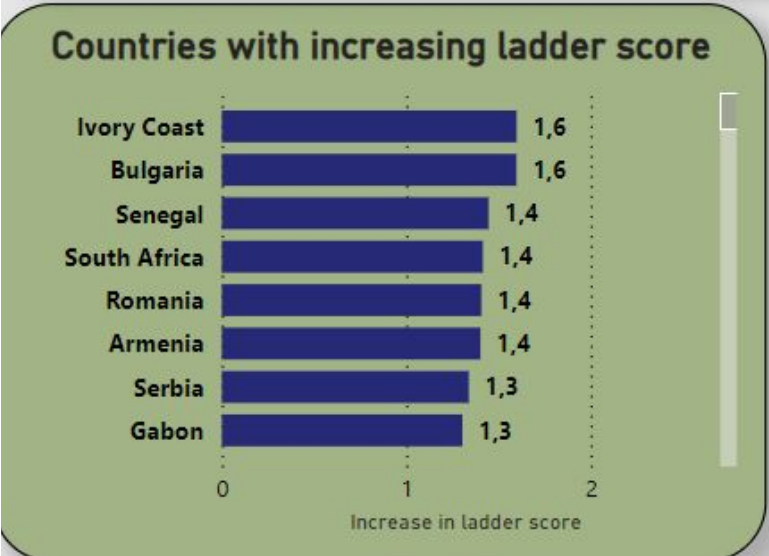
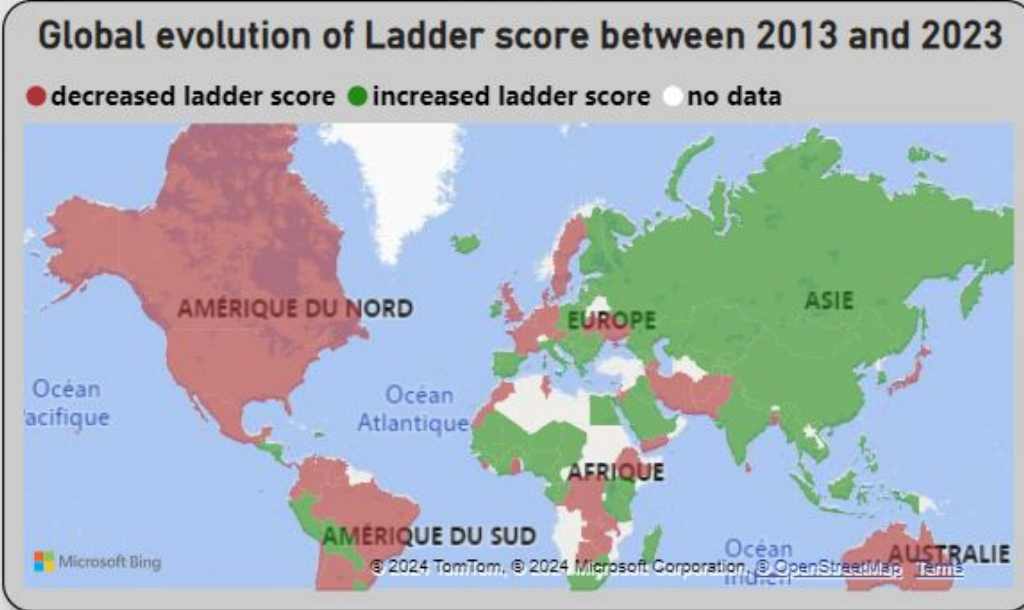
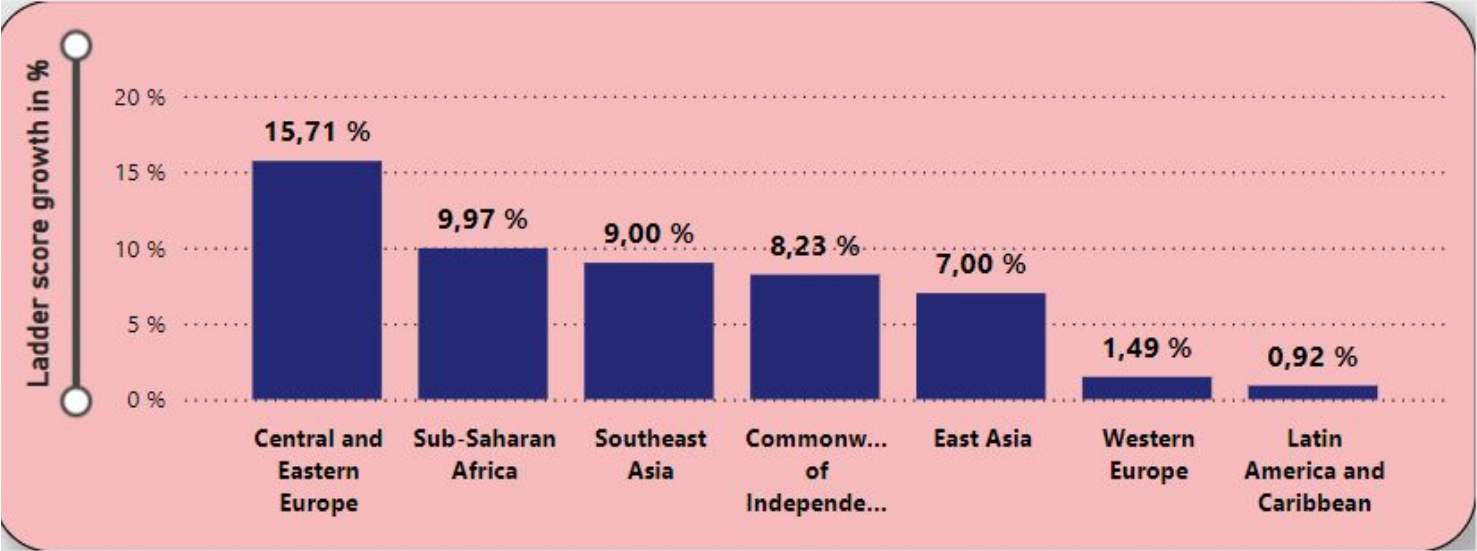


Country name

Tout

Regional indicator

Tout



Correlations between ladder score and different factors

Regional indicator

Central and Eastern Europe

Commonwealth of Independent States

East Asia

Latin America and Caribbean

Country name

☐ Afghanistan

☐ Albania

Average freedom to make lifechoices

0,62

Average healthy life expectancy

0,52

Average social support

1,13

Avg perceived corruption

0,15

Average log GDP per capita

1,38

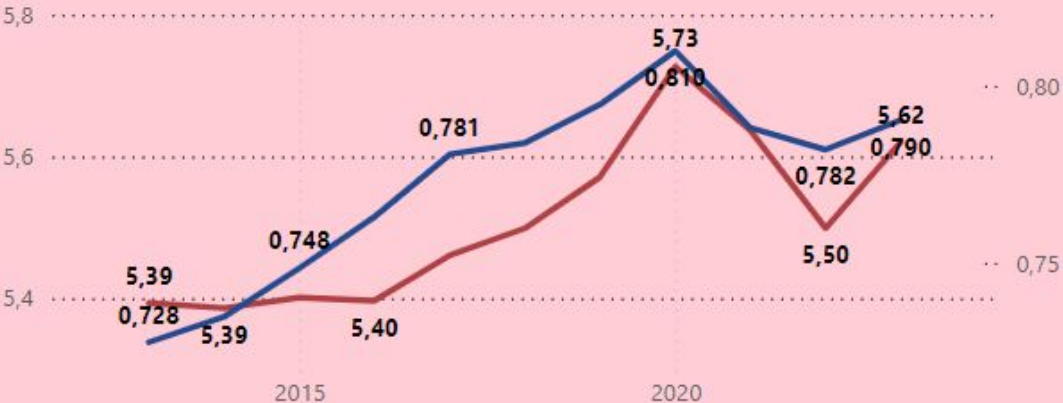
Ladder score and GDP per Capita from 2013 to 2023

● Average Ladder score ● Average Log GDP per capita



Ladder score and Freedom to make life choices from 2013 to 2023

● Average Ladder score ● Average Freedom to make life choices



Correlation between Ladder score and Log GDP per Capita in 2024

● Asia
● Europe
● Latin America and Caribbean
● Middle East and North Africa
● North America and ANZ
● Sub-Saharan Africa



Correlation between ladder score and freedom to make life choices in 2024

● Asia
● Europe
● Latin America and Carib...
● Middle East and North ...
● North America and ANZ
● Sub-Saharan Africa





INTRODUCTION

PRESENTATION

PROBLEMATIQUE

**EXPLORATION DES
DONNEES**

MACHINE LEARNING

CONCLUSION

Objectifs

- Déterminer dans quelle mesure les différents facteurs mesurés permettent de prédire le score de bonheur
- Identifier les facteurs clés du bonheur : déterminer quels indicateurs (PIB, soutien social, espérance de vie, ect.) influencent le plus le score de bonheur d'un pays.
- Prédire les scores futurs : construire des modèles prédictifs pour estimer le niveau de bonheur des pays pour les années futures en se basant sur les données historiques.

Méthodologie

- Trouver le modèle le plus performant
- L'optimiser



INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION



Dataset 2005 - 2023

Country name	year	Ladder score	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices
Afghanistan	2011	3,832	7,581	0,521	51,4	0,496
Afghanistan	2014	3,131	7,671	0,526	52,3	0,509
Afghanistan	2015	3,983	7,654	0,529	52,6	0,389
Afghanistan	2016	4,22	7,65	0,559	52,925	0,523
Afghanistan	2017	2,662	7,648	0,491	53,25	0,427
Afghanistan	2009	4,402	7,509	0,552	50,8	0,679
Afghanistan	2008	3,724	7,35	0,451	50,5	0,718
Afghanistan	2013	3,572	7,68	0,484	52	0,578
Afghanistan	2018	2,694	7,631	0,508	53,575	0,374
Afghanistan	2019	2,375	7,64	0,42	53,9	0,394
Afghanistan	2010	4,758	7,614	0,539	51,1	0,6
Afghanistan	2021	2,436	7,325	0,454	54,55	0,394
Afghanistan	2022	1,281		0,228	54,875	0,368
Afghanistan	2023	1,446		0,368	55,2	0,228
Afghanistan	2012	3,783	7,661	0,521	51,7	0,531
Albania	2023	5,445	9,689	0,691	69,2	0,872
Albania	2016	4,511	9,417	0,638	69,025	0,73
Albania	2007	4,634	9,122	0,821	66,76	0,529
Albania	2009	5,485	9,241	0,833	67,32	0,525
Albania	2010	5,269	9,283	0,733	67,6	0,569
Albania	2011	5,867	9,31	0,759	67,88	0,487
Albania	2012	5,51	9,326	0,785	68,16	0,602

INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES DONNEES

MACHINE LEARNING

CONCLUSION

Remplacement des valeurs nulles sur Python

```
Entrée [13]: colonnes_à_remplacer = ['Log GDP per capita',
    'Social support', 'Healthy life expectancy at birth',
    'Freedom to make life choices', 'Generosity',
    'Perceptions of corruption', 'Positive affect', 'Negative affect']
```

```
Entrée [20]: for col in colonnes_à_remplacer :
    df[col] = df.groupby("Country name")[col].transform(lambda x: x.fillna(x.median()))
```

```
Entrée [24]: df["Perceptions of corruption"].unique()
```

0.473	, 0.487	, 0.573	, 0.409	, 0.337	,
0.373	, 0.36	, 0.358	, 0.905	, 0.902	,
0.748	, 0.922	, 0.671	, 0.744	, 0.716	,
0.74	, 0.659	, 0.609	, 0.67	, 0.643	,
0.594	, 0.709	, 0.68628571	, 0.714	, 0.878	,
0.967	, 0.919	, 0.935	, 0.842	, 0.811	,
0.328	, 0.675	, 0.486	, 0.56	, 0.51225	,
0.929	, 0.637	, 0.587	, 0.65309091	, 0.592	,
0.62	, 0.937	, 0.927	, 0.949	, 0.939	,
0.768	, 0.869	, 0.68466667	, 0.673	, 0.686	,
0.966	, 0.979	, 0.964	, 0.956	, 0.388	,
0.301	, 0.375	, 0.33	, 0.345	, 0.343	,
0.824	, 0.689	, 0.68	, 0.658	, 0.595	,
0.586	, 0.589	, 0.633	, 0.615	, 0.969	,
0.873	, 0.632	, 0.619	, 0.571	, 0.445	,
0.419	, 0.359	, 0.505	, 0.363	, 0.371	,
0.281	, 0.397	, 0.459	, 0.224	, 0.295	,
0.269	, 0.289	, 0.312	, 0.273	, 0.186	,
0.278	, 0.222	, 0.207	, 0.234	, 0.283	,



happiness_python_
inputs



happiness_python_
inputs_prepared



Ladder score Prediction
with Python inputs

INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

Gestion des variables

Rename column 'Life Ladder' to 'Ladder score'

2363

Remove columns col, col_0

2363

+

ADD A NEW STEP

+

ADD A GROUP



Features Handling COPY TO... COPY FROM...

Dataset

Filter

Country name
Reject

OFF

year
Reject

OFF

Ladder score
Target variable

☒

Log GDP per capita
Avg-std rescaling

ON

Social support
Avg-std rescaling

ON

Healthy life expectancy at birth
Avg-std rescaling

ON

Freedom to make life choices
Avg-std rescaling

ON

Generosity
Avg-std rescaling

ON

Perceptions of corruption
Avg-std rescaling

ON

Positive affect
Avg-std rescaling

ON

Negative affect
Avg-std rescaling

ON

INTRODUCTION

PRESENTATION

PROBLEMATIQUE

EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION



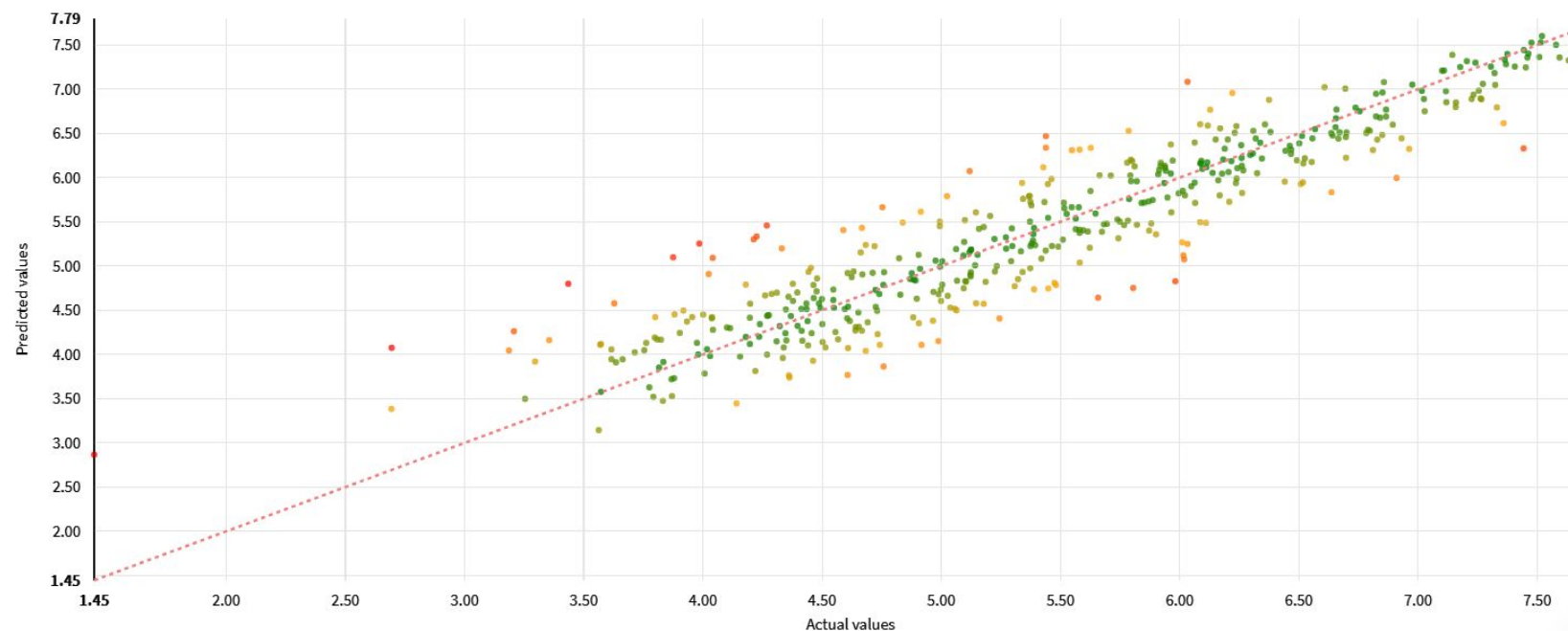
Random forest (test 3 pytho...



0.852



Scatter plot



Après avoir testé différents modèles, nous avons constaté que le modèle de Random Forest est le plus performant en obtenant un R^2 score de 0.852, contre 0.771 pour la Ridge Régression

INTRODUCTION

PRESENTATION

PROBLEMATIQUE

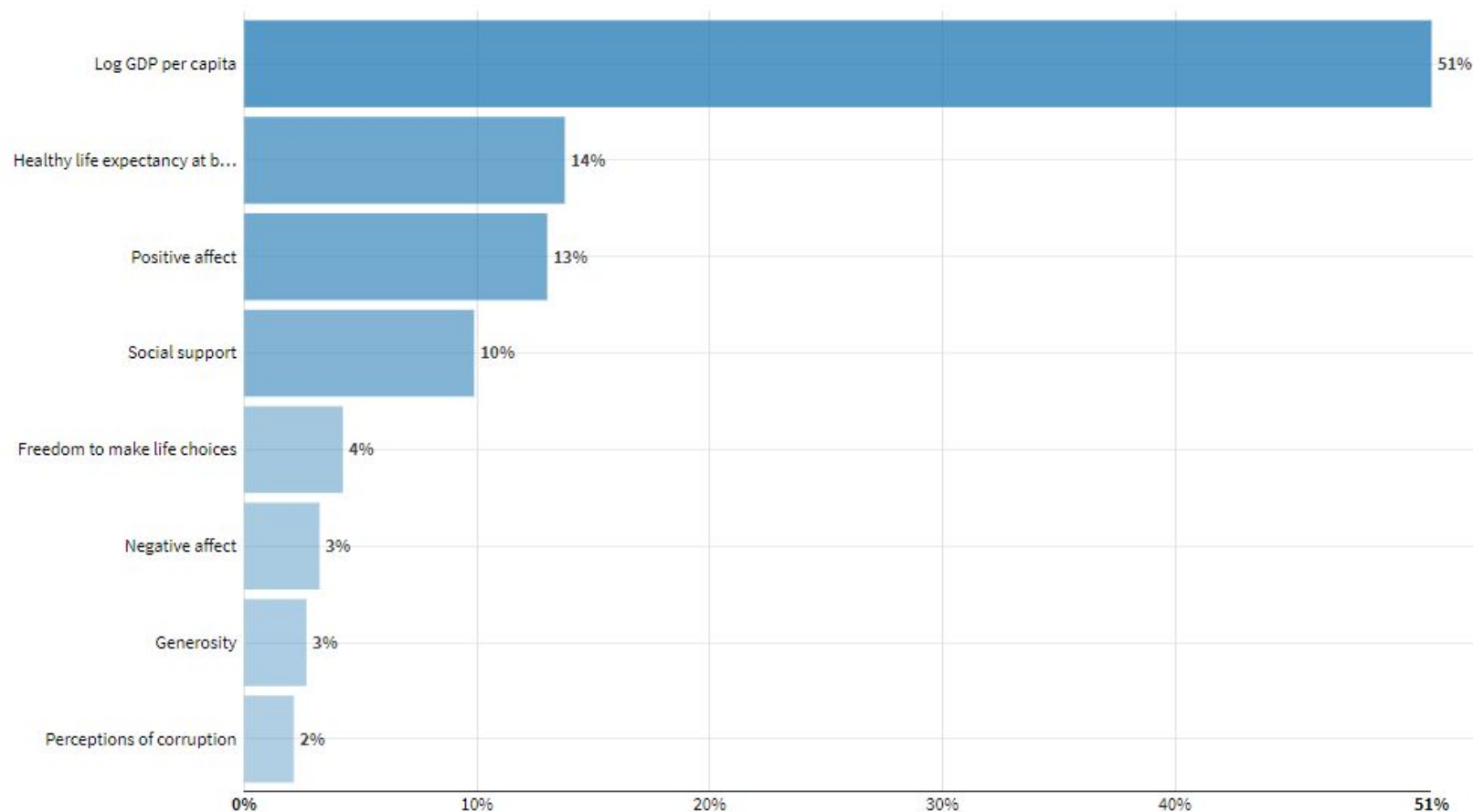
EXPLORATION DES
DONNEES

MACHINE LEARNING

CONCLUSION

Le modèle de Random Forest nous donne le revenu comme principal facteur prédictif

Absolute feature importance



INTRODUCTION

PRESENTATION

PROBLEMATIQUE

**EXPLORATION DES
DONNEES**

MACHINE LEARNING

CONCLUSION



Pour conclure

Les pays les plus "heureux" sont généralement les plus développés économiquement. Les pays dont le niveau de bonheur a le plus augmenté sont également ceux qui ont connu les taux de croissance les plus élevés, en atteste les résultats du machine learning.

Pour aller plus loin

Pour être encore plus précis, nous pouvons rajouter au dataset d'autres facteurs qui peuvent jouer sur le niveau de bonheur dans un pays, comme :

- la sécurité
- l'accès à l'emploi
- les changements climatiques ou catastrophes naturelles
- la stabilité politique



Merci à tous !!!