



**INSTITUT  
de la  
communication**

**UNIVERSITE LUMIERE LYON LUMIERE II**

**Institut de la communication**

**MASTER 1**

**Informatique**

# **Rapport d'étude sur l'état de bonheur dans le monde**

**PROJET SAS**

Sous la supervision de M. Alain Follet

Réalisé par :

- Mame Libasse MBOUP
- Saliou NDAO
- Deffa NDIAYE

Année universitaire 2020-2021

## Table des matières

INTRODUCTION	3
CONTEXTE	4
CONTENU	4
OBJECTIFS	5
ANALYSE DU BONHEUR DANS LE MONDE	5
I. Compréhension des données.....	5
II. Préparation des données .....	6
III. Statistique descriptive .....	9
IV. Régression Multiple .....	14
CONCLUSION	20

## INTRODUCTION

De nos jours, les questions relatives à l'état du bonheur dans le monde constituent un domaine central de recherche pour les sciences sociales, y compris pour la science économique la plus classique. Les gens sont-ils heureux aujourd'hui ? L'étaient-ils plus dans le passé ? comment les personnes évaluent-elles leurs conditions de vie dans des sociétés différentes ? et comment nos propres conditions de vie affectent-elles notre sensation de bonheur ou de satisfaction à l'égard de notre vie présente ? Pour complexe qu'elles soient, ces questions nous touchent personnellement. Ce qui explique tout l'intérêt que ça suscite aujourd'hui, qu'il soit d'ordre politique, économique, social ou autres.

Le Rapport sur le bonheur dans le monde est une publication annuelle du Réseau des solutions de développement durable des Nations Unies. Il contient des articles et des classements du bonheur national basés sur les évaluations des répondants sur leur propre vie, que le rapport met également en corrélation avec divers facteurs de la vie.

Dans les rapports, des experts dans des domaines tels que l'économie, la psychologie, l'analyse d'enquêtes et les statistiques nationales, décrivent comment les mesures du bien-être peuvent être utilisées efficacement pour évaluer les progrès des nations, et d'autres sujets. Chaque rapport est organisé en chapitres qui approfondissent les questions relatives au bonheur, y compris la maladie mentale, les bienfaits objectifs du bonheur, l'importance de l'éthique, les implications politiques et les liens avec l'approche de l'Organisation de Coopération et de Développement Économiques (OCDE) mesurer le bien-être subjectif et d'autres efforts internationaux et nationaux.

Dans ce travail, nous allons exploiter ces rapports afin de comprendre le classement des pays selon leur niveau de bonheur. Nous allons d'abord procéder à la préparation des données, ensuite effectuer une analyse descriptive afin d'observer les tendances et enfin réaliser une régression pour expliquer ces dernières.

## CONTEXTE

Ce projet s'inscrit dans le cadre de notre formation en Statistique et Informatique décisionnelle à l'Université Lumière Lyon 2 -ICOM. Il s'agit de mener une étude sur l'état du bonheur dans le monde avec le langage de programmation SAS. En effet, nous disposons d'un jeu de données, **World Happiness Report**, une étude historique sur les rapports de l'état du bonheur dans le monde de cinq années consécutives (2015 à 2019). Ces derniers classent les pays selon leur niveau de bonheur. Le rapport continue de gagner une reconnaissance mondiale alors que les gouvernements, les organisations et la société civile utilisent de plus en plus des indicateurs de bonheur pour éclairer leurs décisions politiques. Des experts de premier plan dans tous les domaines - économie, psychologie, analyse d'enquêtes, statistiques nationales, santé, politique publique et plus - décrivent comment les mesures du bien-être peuvent être utilisées efficacement pour évaluer les progrès des nations.

## CONTENU

Les scores de bonheur et les classements utilisent les données du Gallup World Poll. Les scores sont basés sur les réponses à la principale question d'évaluation de la vie posée dans le sondage. Cette question, connue sous le nom d'échelle de Cantril, demande aux répondants de penser à une échelle avec la meilleure vie possible pour eux étant un 10 et la pire vie possible étant un 0 et d'évaluer leur propre vie actuelle sur cette échelle. Les scores proviennent d'échantillons représentatifs au niveau national pour les années 2015-2019 et utilisent les poids Gallup pour rendre les estimations représentatives. Les colonnes qui suivent le score de bonheur estiment dans quelle mesure chacun des six facteurs - production économique, soutien social, espérance de vie, liberté, absence de corruption et générosité - contribue à rendre les évaluations de vie plus élevées dans chaque pays qu'elles ne le sont dans la Dystopie, un pays hypothétique dont les valeurs sont égales aux moyennes nationales les plus basses du monde pour chacun des six facteurs. Ils n'ont aucun impact sur le score total rapporté pour chaque pays, mais ils expliquent pourquoi certains pays se classent plus haut que d'autres.

## OBJECTIFS

Dans l'ensemble, nous avons quelques questions que nous voulions aborder avec notre analyse des données :

1. Quels sont les facteurs contribuant au bonheur ? Autrement dit les variables qui ont contribué le plus au bonheur général d'un pays ou qui contribuent le plus au score de bonheur du pays. Un changement significatif de ces variables au sein d'un pays aura-t-il un fort ressenti sur la société ?
2. Étudier la corrélation entre les variables.
3. Quels pays se classent le mieux en termes de bonheur global et de chacun de ces facteurs.
4. Comment les scores des pays ont évolué en fonction des années (2015, 2016..., 2019) ?
5. Existe-t-il une augmentation ou diminution du bonheur d'un pays au fil des années ?

## ANALYSE DU BONHEUR DANS LE MONDE

### I. Compréhension des données

Nous avons à notre disposition les rapports d'état du bonheur dans le monde de 2015 à 2019. Pour comprendre nos données, il nous a fallu faire une analyse simple de nos rapports. Ces derniers contiennent des scores de bonheur et des classements qui utilisent les données du Gallup World Poll. Les scores de ces ensembles de données sont basés sur les principales questions d'évaluation de la vie posées dans le sondage.

Notre jeu de données contient 13 variables (dont 2 catégorielles, 11 quantitatives) et 155 observations.

#### Variables qualitatives

- **Pays** : variable nominale
- **Région** : variable nominale

#### Variables Quantitatives

- **Happiness\_Rank** : rang au classement du bonheur (de 1 à 156)
- **Happiness\_Score** : score de bonheur (échelle de 1 à 10)
- **Lower\_Confidence\_Interval, Upper\_Confidence\_Interval** : prédit les intervalles de confiance du score de bonheur d'un pays

- **Economy\_GDP\_per\_Capita\_** : nous renseigne sur la performance économique d'un pays (PIB).
- **Famille** : Support social
- **Health\_Life\_Expectancy\_** : renseigne sur l'espérance de vie en se basant sur l'accès et la qualité des soins médicaux.
- **Freedom** : liberté que les gens sentent dans leurs pays
- **Trust\_Government\_Corruption\_** : confiance accordée au gouvernement
- **Generosity** : renseigne sur la générosité des gens dans un pays données.
- **Dystopia\_Residual** : C'est un pays imaginaire qui compte les personnes les moins heureuses du monde. Cette variable est utilisée pour comparer avec d'autres pays et évaluer leur niveau de bonheur.

Nous avons considéré la variable **score de bonheur** comme étant notre variable de classe.

## II. Préparation des données

### II.1 Nettoyage des données

Comme à l'accoutumée, cette étape consiste à préparer le jeu de données pour les futurs traitements. Il s'agira de voir combien de nos données manquent, si elles sont cohérentes, si elles présentent des valeurs aberrantes, etc. Pour le cas des valeurs manquantes, nous avons parcouru les cinq fichiers de notre ensemble de données et constaté que nous n'avions aucune valeur manquante. Nous avons comparé les fichiers qui nous ont été fournis pour nous assurer que les données étaient cohérentes. En examinant de plus près l'ensemble de données, nous avons constaté que peu de modifications étaient nécessaires, car les données fournies étaient lisibles, complètes, cohérentes et correctes. En gros, cet ensemble de données était très propre.

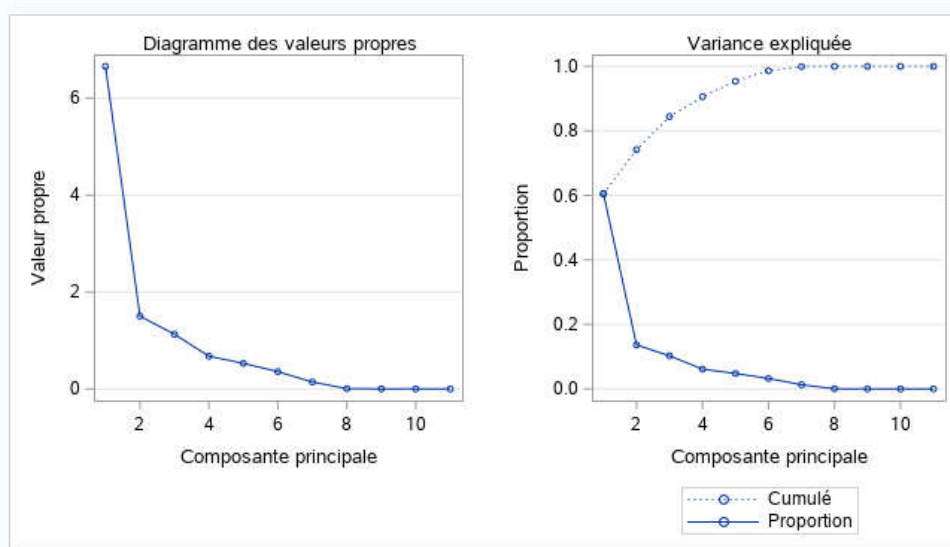
### II.1 Réduction et intégration de données.

Toujours dans le cadre de la préparation des données, nous avons cherché à recenser les variables les plus informatives ainsi que leur corrélation.

Nous avons effectué **une analyse en composante principale** sur nos variables quantitatives afin de déceler celles qui sont les plus importantes, puis nous avons regardé leur corrélation afin de voir les types de relations entre ces dernières. Nous avons cherché par la suite s'il y avait des variables portant la même information(redondance) que nous pourrions supprimer ou en garder une.

Nous illustrons nos propos avec les figures ci-dessous.

Vecteurs propres											
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10	Prin11
Happiness_Rank	-0.382419	0.083163	-0.088971	-0.033312	-0.038720	0.000543	0.058226	0.910980	0.058243	0.000022	0.000000
Happiness_Score	0.383788	-0.071475	0.094686	0.029517	0.034284	0.016362	-0.043285	0.181452	0.008199	-0.367455	-0.818351
Lower_Confidence_Interval	0.383913	-0.068899	0.090124	0.028561	0.036031	0.019406	-0.023337	0.224454	-0.693697	-0.369685	0.410453
Upper_Confidence_Interval	0.383259	-0.074205	0.099195	0.030452	0.032484	0.013269	-0.063391	0.137824	0.717219	-0.364819	0.408324
Economy__GDP_per_Capita__	0.322401	-0.071727	-0.430301	-0.088169	0.159341	-0.108704	-0.693001	0.149366	-0.009196	0.398093	-0.000000
Family	0.300880	-0.077738	-0.256648	0.200857	-0.512139	0.649043	0.218973	0.072878	0.012829	0.257356	-0.000000
Health__Life_Expectancy__	0.312652	-0.052377	-0.382721	0.072764	0.362446	-0.352922	0.660901	0.073971	0.024221	0.221281	-0.000000
Freedom	0.239109	0.417030	0.164887	-0.108865	-0.623339	-0.564948	0.015313	0.048950	-0.000367	0.140420	-0.000000
Trust__Government_Corruption__	0.174806	0.484351	0.143188	-0.708137	0.275613	0.336938	0.106826	0.021105	0.008117	0.107113	-0.000000
Generosity	0.070107	0.612047	0.218279	0.655585	0.323929	0.093885	-0.105520	0.038803	0.000303	0.129081	-0.000000
Dystopia_Residual	0.165283	-0.432078	0.687510	0.010588	0.080533	0.006656	0.049026	0.174835	0.004510	0.523153	-0.000000



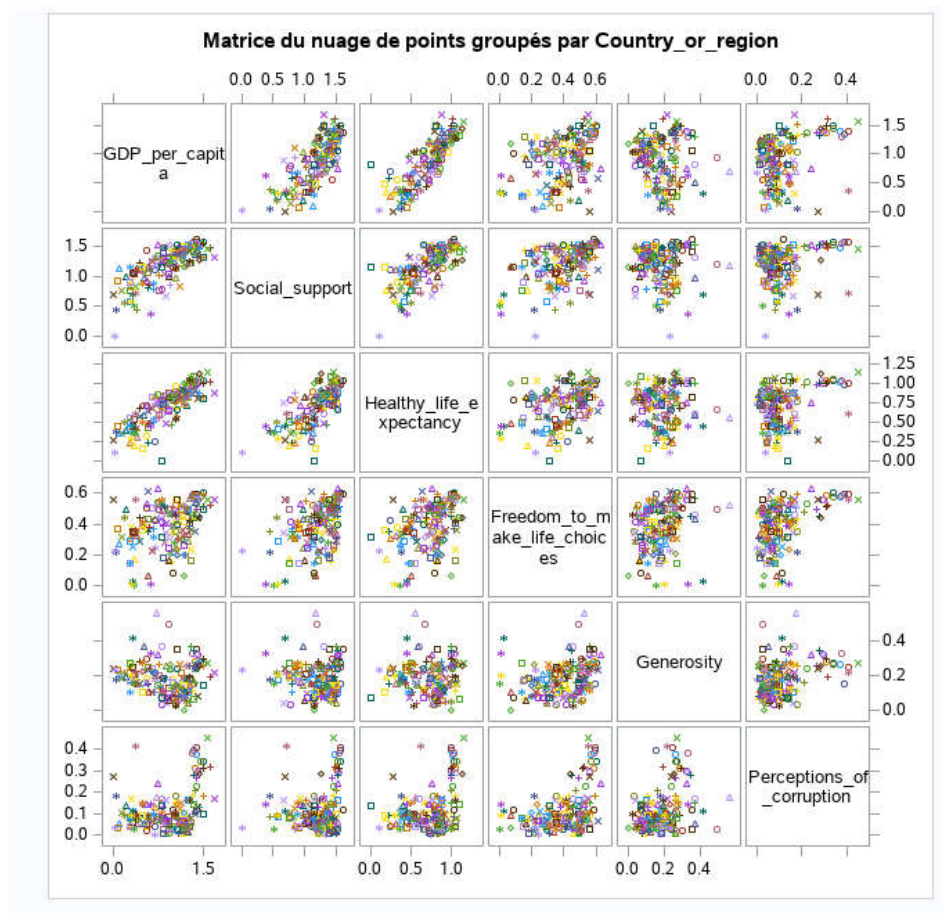
**Figure 1 : Résultats de l'ACP**

Matrice de corrélation											
	Happiness_Rank	Happiness_Score	Lower_Confidence_Interval	Upper_Confidence_Interval	Economy__GDP_per_Capita__	Family	Health__Life_Expectancy__	Freedom	Trust__Government_Corruption__	Generosity	Dystopia_Residual
Happiness_Rank	1.0000	-0.9957	-0.9949	-0.9955	-0.7930	-0.7333	-0.7680	-0.5572	-0.3871	-0.1454	
Happiness_Score	-0.9957	1.0000	0.9995	0.9995	0.7903	0.7393	0.7854	0.5968	0.4020	0.1598	
Lower_Confidence_Interval	-0.9949	0.9995	1.0000	0.9979	0.7907	0.7415	0.7892	0.5969	0.4048	0.1580	
Upper_Confidence_Interval	-0.9955	0.9995	0.9979	1.0000	0.7891	0.7362	0.7808	0.5962	0.3988	0.1555	
Economy__GDP_per_Capita__	-0.7930	0.7903	0.7907	0.7891	1.0000	0.8698	0.8371	0.3623	0.2942	-0.2555	
Family	-0.7333	0.7393	0.7415	0.7362	0.8698	1.0000	0.5894	0.4502	0.2136	0.0896	
Health__Life_Expectancy__	-0.7680	0.7854	0.7892	0.7808	0.8371	0.5894	1.0000	0.3412	0.2496	0.0780	
Freedom	-0.5572	0.5968	0.5969	0.5962	0.3623	0.4502	0.3412	1.0000	0.5021	0.3518	
Trust__Government_Corruption__	-0.3871	0.4020	0.4048	0.3988	0.2942	0.2136	0.2496	0.5021	1.0000	0.3059	
Generosity	-0.1454	0.1598	0.1580	0.1555	-0.2555	0.0896	0.0780	0.3518	0.3059	1.0000	
Dystopia_Residual	-0.5420	0.5437	0.5388	0.5482	0.0896	0.1197	0.1009	0.0916	-0.0029	-0.1330	

**Figure 2 : Matrice de corrélation des variables quantitatives**

D'après cette matrice des corrélations, nous pouvons voir les variables qui sont fortement corrélées avec le score de bonheur autrement dit les facteurs contribuant au bonheur des pays.

Nous avons conclu que nous ne pouvions pas supprimer ou intégrer des variables en raison du degré de classification de chaque variable.



***Figure 3 : Matrice des nuages de points***

Aux vues de la matrice des nuages de points nous constatons que :

- La variable GDP\_per\_capita (PIB) est corrélée avec les variables social\_support et healthy\_life\_expectancy qui sont elles aussi corrélées entre elles.
- La variable Freedom, Generosity et Perceptions\_of\_corruption ne sont corrélées avec aucune autre variable.



## II.1 Transformation des données

Étant donné que nos données étaient déjà réparties pour chaque année et étaient propres, nous les avons trouvées faciles à comprendre et nous avons permis de réduire le temps passé au prétraitement et transformation. En résumé, aucune transformation n'a été faite, et toutes les variables sont pertinentes pour nos analyses. Cependant, dans les analyses qui suivent, nous allons essayer de mettre en évidence les variables les plus pertinentes pour le score de bonheur d'un pays.

## III. Statistique descriptive

Nous rappelons que le but principal de notre étude est d'explorer les facteurs contribuant le plus au bonheur d'un pays. Ce qui permet de décrire ce qui s'est passé avec nos données de 2015 à 2019. Pour atteindre nos objectifs, nous allons dégager quelques axes d'analyse grâce à la statistique descriptive.

Variable	Moyenne	Ec-type	Minimum	Maximum	N
Overall_rank	78.5000000	45.1774280	1.0000000	156.0000000	156
Score	5.4070962	1.1131199	2.8530000	7.7690000	156
GDP_per_capita	0.9051474	0.3983895	0	1.6840000	156
Social_support	1.2088141	0.2991914	0	1.6240000	156
Healthy_life_expectancy	0.7252436	0.2421240	0	1.1410000	156
Freedom_to_make_life_choices	0.3925705	0.1432895	0	0.6310000	156
Generosity	0.1848462	0.0952544	0	0.5660000	156
Perceptions_of_corruption	0.1106026	0.0945378	0	0.4530000	156

### ❖ Top 10 des pays de faible/ forte économie

D'après le tableau ci-dessous on constate que l'économie de Somalie est quasi nulle, cependant ce dernier n'a pas le plus faible score de bonheur, mais les pays constituant ce tableau ont un score de bonheur faible comparé au pays de forte économie. Donc on peut présumer que la variable GDP\_per\_capita pèse beaucoup sur le classement des niveaux de bonheur. Néanmoins, il existe d'autres variables ou combinaison de variables pesant lourd sur le classement.

	Country_or_region	Score	GDP_per_capita
1	Somalia	4.982	0
2	Central Africa	3.083	0.024
3	Congo (Kinshas	4.245	0.069
4	Liberia	3.495	0.076
5	Burundi	2.905	0.091
6	Niger	4.166	0.131
7	Malawi	3.587	0.186
8	Mozambique	4.417	0.198
9	Sierra Leone	4.571	0.256
10	Togo	3.999	0.259

**Tableau 1 : pays avec faible économie**

	Country_or_region	Score	GDP_per_capita
1	United Arab Em	6.774	2.096
2	Qatar	6.374	1.649
3	Luxembourg	6.91	1.576
4	Singapore	6.343	1.529
5	Kuwait	6.083	1.474
6	Norway	7.594	1.456
7	Ireland	6.977	1.448
8	Switzerland	7.487	1.42
9	Hong Kong	5.43	1.405
10	United States	6.886	1.398

**Tableau 2 : pays avec forte économie**

Si nous comparons les pays à une même date, nous constatons automatiquement que les pays disposant d'une croissance économique plus élevée affiche des niveaux de satisfaction de vie plus élevée.

#### ❖ Top 10 des pays avec un fort / faible Social\_suport

Tout comme la variable GDP\_per\_capita, la variable Social\_support aurait une forte influence sur le classement du niveau de bonheur. En comparant les deux tableaux on voit nettement que les pays avec un support social élevé sont des pays avec un haut niveau de bonheur.

	Country_or_region	Score	Social_support
1	Iceland	7.494	1.624
2	Finland	7.769	1.587
3	Norway	7.554	1.582
4	Denmark	7.6	1.573
5	New Zealand	7.307	1.557
6	Ireland	7.021	1.553
7	Australia	7.228	1.548
8	Turkmenistan	5.247	1.538
9	United Kingdom	7.054	1.538
10	Mongolia	5.285	1.531

**Tableau 3 : pays avec un fort support social**

On constate que le support social en Afrique Central est quasi nul et son score de bonheur est très faible, ce qui confirme nos analyses faites précédemment.

	Country_or_region	Score	Social_support
1	Central Africa	3.083	0
2	Syria	3.462	0.378
3	Benin	4.883	0.437
4	Burundi	3.775	0.447
5	Afghanistan	3.203	0.517
6	Malawi	3.41	0.56
7	Togo	4.085	0.572
8	South Sudan	2.853	0.575
9	Georgia	4.519	0.666
10	Haiti	3.597	0.688

**Tableau 4 : pays avec un faible support social**

❖ **Top 10 des pays avec un fort / faible Health\_Life\_Expectancy\_**

	Country_or_region	Score	Healthy_life_expectancy
1	Swaziland	4.212	0
2	Central Africa	3.083	0.105
3	Lesotho	3.802	0.168
4	Chad	4.35	0.192
5	Ivory Coast	4.944	0.232
6	Sierra Leone	4.374	0.242
7	Nigeria	5.265	0.245
8	Somalia	4.668	0.268
9	South Sudan	2.853	0.295
10	Mali	4.39	0.308

**Tableau 5 : Pays avec faible taux de Health\_Life\_Expectancy\_**

Dans la même logique on remarque que les pays dont les soins médicaux ne sont pas très accessibles ont un niveau de bonheur plutôt faible. Pour le cas du Nigeria cela peut s'expliquer du fait de la forte population nigériane.

	Country_or_region	Score	Healthy_life_expectancy
1	Singapore	6.262	1.141
2	Hong Kong	5.43	1.122
3	Japan	5.886	1.088
4	Spain	6.354	1.062
5	Switzerland	7.48	1.052
6	France	6.592	1.045
7	Northern Cypru	5.718	1.042
8	Cyprus	6.046	1.042
9	Italy	6.223	1.039
10	Canada	7.278	1.039

**Tableau 6 : Pays avec fort taux de Health\_Life\_Expectancy\_**

On constate que l'accès aux soins ainsi que la qualité de ces derniers contribuent largement sur le niveau de bonheur d'un pays. On peut citer l'exemple de la Suisse et du Canada qui sont des pays développés donc forcément avec des services sanitaires à la hauteur qui vont influencer sur l'échelle de satisfaction.

❖ **Top 10 des pays avec un fort / faible Freedom**

	Country_or_region	Score	Freedom_to_make_life_choices
1	Uzbekistan	6.174	0.631
2	Cambodia	4.7	0.609
3	Norway	7.554	0.603
4	United Arab Em	6.825	0.598
5	Finland	7.769	0.596
6	Denmark	7.6	0.592
7	Iceland	7.494	0.591
8	New Zealand	7.307	0.585
9	Canada	7.278	0.584
10	Sweden	7.343	0.574

**Tableau 7 : Pays avec un fort pourcentage de Freedom**

On retrouve les mêmes nations, représentatives sur toutes les formes de liberté telles que la Suède, le Canada et la Nouvelle Zélande ce qui semblerait indiquer que les pays avec une liberté de choix de vie avancée ont une forte échelle de bonheur.

	Country_or_region	Score	Freedom_to_make_life_choices
1	Afghanistan	3.203	0
2	South Sudan	2.853	0.01
3	Syria	3.462	0.013
4	Haiti	3.597	0.026
5	Mauritania	4.49	0.066
6	Greece	5.287	0.067
7	Algeria	5.211	0.086
8	Comoros	3.973	0.142
9	Yemen	3.38	0.143
10	Madagascar	3.933	0.148

**Tableau 8 : Pays avec un faible pourcentage de Freedom**

L'Afghanistan a un pourcentage totalement nul, c'est un pays instable. C'est le cas également de l'Algérie, du Yémen et de la Syrie ; avec leur faible taux de liberté ils ont des scores de niveau de bonheur assez faibles. Ce qui traduit la forte influence de la liberté sur le niveau moyen de bonheur.

#### ❖ Evolution du score de bonheur au cours de la période 2015 à 2019

Nous constatons que les scores augmentent et baissent selon les années et les pays. Ces variations sont certainement liées aux fluctuations des autres variables impactant aussi le niveau du score de bonheur.

	Country	Evolution2015_2016	Evolution2016_2017	Evolution2017_2018	Evolution2018_2019
1	Lesotho	Baisse	Baisse	Baisse	Baisse
2	Czech Republi	Augmentation	Baisse	Baisse	Baisse
3	Dominican Rep	Augmentation	Baisse	Baisse	Baisse
4	Suriname	Baisse	Baisse	Baisse	Baisse
5	United Arab E	Baisse	Baisse	Baisse	Baisse
6	Palestinian T	Augmentation	Baisse	Baisse	Baisse
7	Congo (Kinsha	Baisse	Baisse	Baisse	Baisse
8	Mozambique	Baisse	Baisse	Baisse	Baisse
9	Bosnia and He	Augmentation	Baisse	Baisse	Baisse
10	Somaliland re	Baisse	Baisse	Baisse	Baisse

**Tableau 9 : Evolution du bonheur au cours de la période 2015 à 2019**

En résumé, les variables qui semblent influencer sur le niveau de bonheur sont Freedom, Social\_support, Economy\_GDP\_per\_Capita\_ et Health\_Life\_Expectancy\_.

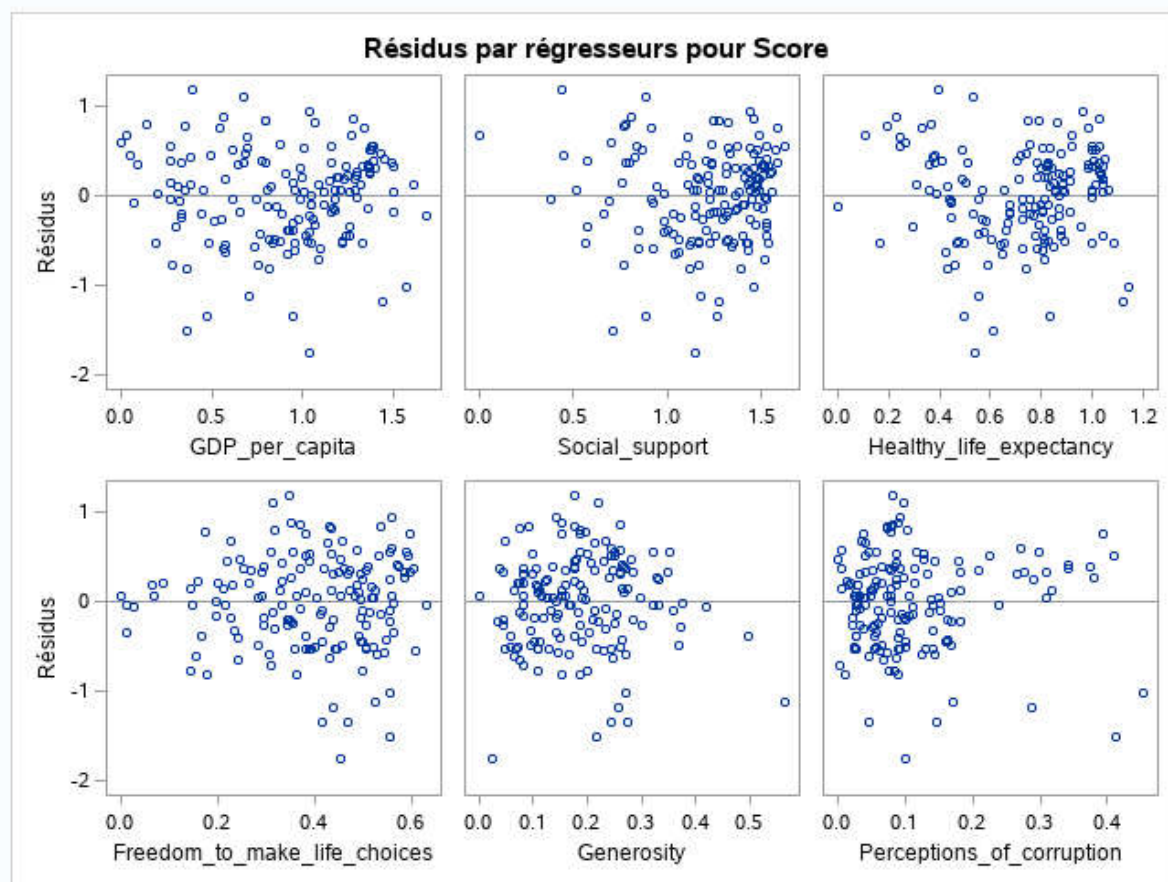
Pour vérifier la véracité de nos hypothèses, nous allons effectuer une régression multiple dans les lignes qui suivent.

#### IV. Régression Multiple

En statistique, la régression linéaire multiple est une méthode de régression mathématique étendant la régression linéaire simple pour décrire les variations d'une variable endogène associée aux variations de plusieurs variables exogènes (Wikipédia). Il s'agira dans notre cas, d'appliquer cette méthode afin d'étudier le niveau de bonheur d'un pays en fonction de son économie, de l'accès et de la qualité des soins, du support familial et du degré de liberté.

Ci-dessous, nous présentons les résultats de la régression.

## ❖ Analyse des résidus



**Figure 4: Représentation des résidus**

Afin de pouvoir interpréter les résultats de la régression, nous vérifions les hypothèses suivantes:

- **Moyenne des résidus est nulle :**

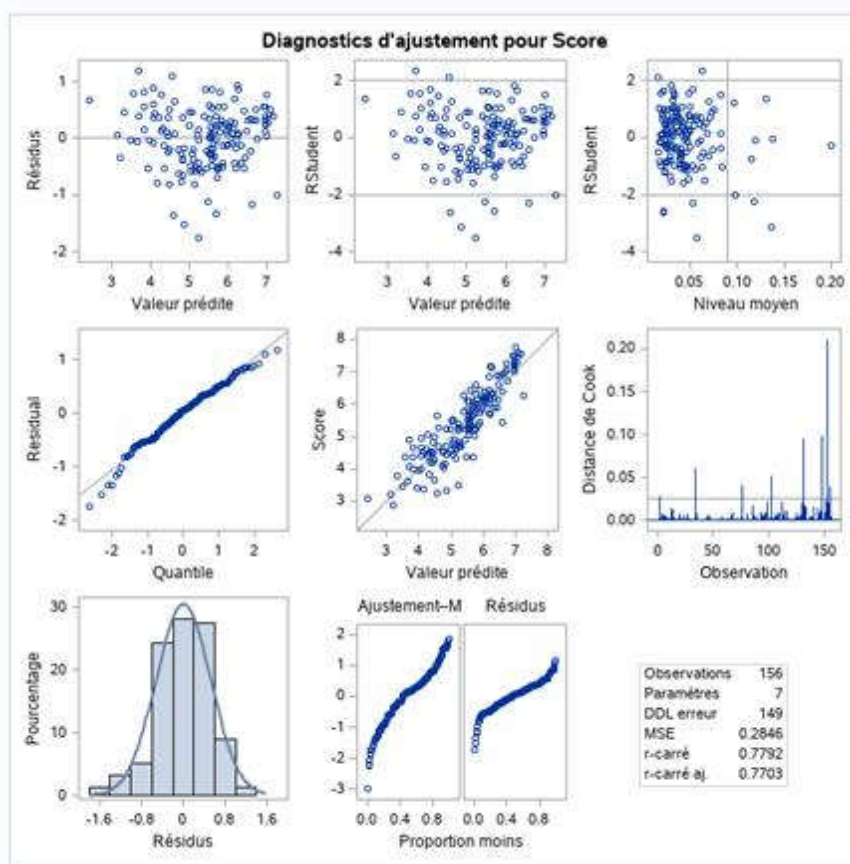
D'après les graphiques des résidus en fonction de nos variables explicatives ci-dessus, nous pouvons observer que les points sont uniformément répartis autour de 0. Ce qui nous permet de supposer que la moyenne est nulle.

- **Variances constantes :**

En examinant les graphiques nous remarquons que la variance de l'erreur des variables exogènes est la même. Ce qui nous permet de supposer l'homoscédasticité des variances.

- **Normalité résidus**

Les résidus se sont alignés sur la droite donc nous pouvons supposer la normalité.



**Figure 5:** Ajustement du score

A présent nous pouvons interpréter les résultats de notre régression.

#### ❖ Significativité globale du modèle

Analyse de variance					
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Modèle	6	149.63884	24.93981	87.62	< .0001
Erreur	149	42.41171	0.28464		
Total sommes corrigées	155	192.05056			

**Tableau 9:** Significativité globale du modèle

Nous constatons que le modèle est globalement significatif puisque sa probabilité critique (p-value) est inférieure à 0.05.



### ❖ Interprétation du coefficient de détermination (R carré)

Le R carré mesure la significativité globale du modèle explicatif. Ce coefficient rapporte la variance expliquée à la variance totale : plus il est voisin de l'unité, meilleure est l'explication.

Ainsi nous constatons que 77% des fluctuations de la probabilité d'un pays d'être heureux sont expliquées par les variables : GDP\_per\_Capita, Social\_support, healthy\_life\_expectancy et Freedom\_to\_make\_life\_choices. Les 23% restants représentent notre taux d'erreur.

Root MSE	0.53352	R carré	0.7792
Moyenne dépendante	5.40710	R car. ajust.	0.7703
Coeff Var	9.86701		

**Tableau 10: coefficient de détermination (R carré)**

### ❖ Interprétation des résultats

Résultats estimés des paramètres					
Variable	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr >  t
Intercept	1	1.79522	0.21107	8.51	<.0001
GDP_per_capita	1	0.77537	0.21823	3.55	0.0005
Social_support	1	1.12419	0.23690	4.75	<.0001
Healthy_life_expectancy	1	1.07814	0.33454	3.22	0.0016
Freedom_to_make_life_choices	1	1.45483	0.37534	3.88	0.0002
Generosity	1	0.48978	0.49775	0.98	0.3267
Perceptions_of_corruption	1	0.97228	0.54236	1.79	0.0751

**Tableau 11: Estimations des paramètres**

A travers le tableau, nous confirmons les analyses faites dans la partie statistique descriptive. Nous constatons que les variables **GDP\_per\_Capita**, **Social\_support**, **healthy\_life\_expectancy** et **Freedom\_to\_make\_life\_choices** influencent significativement le score de bonheur car ayant une p-value inférieure au risque alpha (0.05). Ces variables sont des facteurs positifs du score de bonheur. Plus leurs valeurs augmentent, plus le niveau de bonheur augmente. Ce qui va à l'encontre de ce que l'on observe pour les variables **Generosity** et **Perception Of Corruption**, p-value supérieure au risque alpha (0.05). Ces dernières ont un taux de contribution négligeable.

Country	Evolution2018_2019	PIB2018_2019	Family2018_2019	Health2018_2019	Freedom2018_2019
North Cyprus	Baisse	Baisse	Baisse	Baisse	Baisse
Taiwan	Baisse	Baisse	Baisse	Baisse	Baisse
Oman	Baisse	Baisse	Baisse	Baisse	Baisse
Congo (Brazza	Baisse	Baisse	Baisse	Baisse	Baisse
Afghanistan	Baisse	Augmentation	Baisse	Augmentation	Baisse
Albania	Augmentation	Augmentation	Augmentation	Augmentation	Baisse
Algeria	Baisse	Augmentation	Augmentation	Augmentation	Augmentation
Angola	Baisse	Baisse	Baisse	Baisse	Baisse
Argentina	Baisse	Augmentation	Baisse	Augmentation	Baisse
Armenia	Augmentation	Augmentation	Augmentation	Augmentation	Augmentation

**Tableau 12:** Evolution du score en fonction des facteurs significatifs entre 2018 et 2019

Le tableau suivant est une confirmation que notre modèle est performant bien qu'étant pas fiable à 100% car nous avons un taux d'erreur de 23% ce qui pourrait expliquer le fait que pour l'Algérie nous voyons que son niveau de bonheur est en baisse malgré l'augmentation des tous les autres facteurs influençant le score.

### ❖ Ajustement du modèle

Nous allons ainsi refaire le modèle en supprimant les variables **Generosity** et **Perception\_of\_corruption** qui n'ont pas trop d'effet sur l'état de bonheur d'un pays.

Analyse de variance					
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Modèle	4	148.05798	37.01450	127.05	< .0001
Erreur	151	43.99257	0.29134		
Total sommes corrigées	155	192.05056			

Root MSE	0.53976	R carré	0.7709
Moyenne dépendante	5.40710	R car. ajust.	0.7649
Coeff Var	9.98245		

Résultats estimés des paramètres					
Variable	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr >  t
Intercept	1	1.89210	0.19936	9.49	< .0001
GDP_per_capita	1	0.81055	0.21645	3.74	0.0003
Social_support	1	1.01661	0.23475	4.33	< .0001
Healthy_life_expectancy	1	1.14145	0.33731	3.38	0.0009
Freedom_to_make_life_choices	1	1.84582	0.34039	5.42	< .0001

**Tableau 13:** Résultats du modèle ajusté

Les résultats obtenus avec ce nouveau modèle sont pratiquement les mêmes que précédemment donc le modèle n'a pas gagné en performance. Néanmoins, le déploiement du modèle sera plus facile du fait du nombre réduit de variables.

### ❖ Croisement des variables

Dans cette partie, nous effectuons à nouveau la régression multiple en mettant en évidence l'interaction de certaines variables.

Modèle des moindres carrés (Aucune sélection)

Analyse de variance					
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Modèle	10	157.17772	15.71777	65.35	<.0001
Erreur	145	34.87284	0.24050		
Total sommes corrigées	155	192.05056			

Racine MSE	0.49041
Moyenne dépendante	5.40710
R carré	0.8184
R car. ajust.	0.8059
AIC	-53.71105
AICC	-51.52923
SBC	-178.16263

Résultats estimés des paramètres					
Paramètre	DDL	Estimation	Erreur type	Valeur du test t	Pr >  t
Intercept	1	4.002260	0.444249	9.01	<.0001
GDP_per_capita	1	0.451855	0.981643	0.46	0.6460
Social_support	1	-0.924859	0.671528	-1.38	0.1706
Healthy_life_expecta	1	-2.237360	1.717353	-1.30	0.1947
Freedom_to_make_life	1	1.558112	1.337752	1.16	0.2460
GDP_per_c*Socia_sup	1	-0.662503	0.857997	-0.77	0.4413
GDP_per_c*Healthy_li	1	0.522473	0.771400	0.68	0.4993
GDP_per_c*Freedom_to	1	1.515853	1.799082	0.84	0.4009
Social_su*Healthy_li	1	3.676089	1.539861	2.39	0.0183
Social_su*Freedom_to	1	1.259635	1.573145	0.80	0.4246
Healthy_l*Freedom_to	1	-3.628564	3.234758	-1.12	0.2638

**Tableau 14: Croisement des variables**

En croisant certaines variables entre elles nous obtenons les résultats ci-dessus. Le modèle est globalement très significatif avec un taux d'erreur réduit à 20%. Cependant seul l'interaction entre la Social\_support et Healthy\_life\_expectancy a un impact significatif sur le niveau de bonheur.

## CONCLUSION

Il résulte de ce qui précède que l'objectif de ce projet a été d'étudier l'état de bonheur dans le monde avec le langage de programmation SAS. En effet le niveau de bonheur d'un pays est mesuré par la variable score ; plus elle est importante plus le pays est heureux. Cependant plusieurs critères(variables) influent sur le score. Nous avons pu les recenser en suivant une démarche méthodologique cohérente.

Nous avons réalisé une étude statistique afin de sortir différentes hypothèses. Il est dès lors question de vérifier la véracité de ces dernières en faisant recours aux algorithmes d'apprentissage supervisés, notamment la régression multiple. Ce qui nous a permis de voir les variables qui contribuaient le plus au niveau de bonheur d'un pays donné c'est à dire l'économie, l'accès et la qualité des soins, la liberté, le support social...Nous avons pu montrer l'évolution des scores de bonheur de chaque pays en fonction des années, comment et dans quel sens les facteurs(variables) retenus pouvaient jouer sur cette variation.

Bien que nous ayons effectué une analyse complète à l'aide de la régression, on aurait pu être beaucoup plus rigoureux sur les hypothèses de la régression, en vérifiant par exemple l'existence d'un éventuel phénomène de multicolinéarité (c'est à dire gérer les variables exogènes qui ont une forte corrélation entre elles en les combinant ou en choisissant celle de forte corrélation avec le score) ou en explorant la distance de Cook (Figure 5) afin d'examiner les points influents autrement dit les observations qui peuvent avoir un impact sur l'ajustement de la régression.

Source : <https://www.kaggle.com/unsdsn/world-happiness?select=2019.csv>