

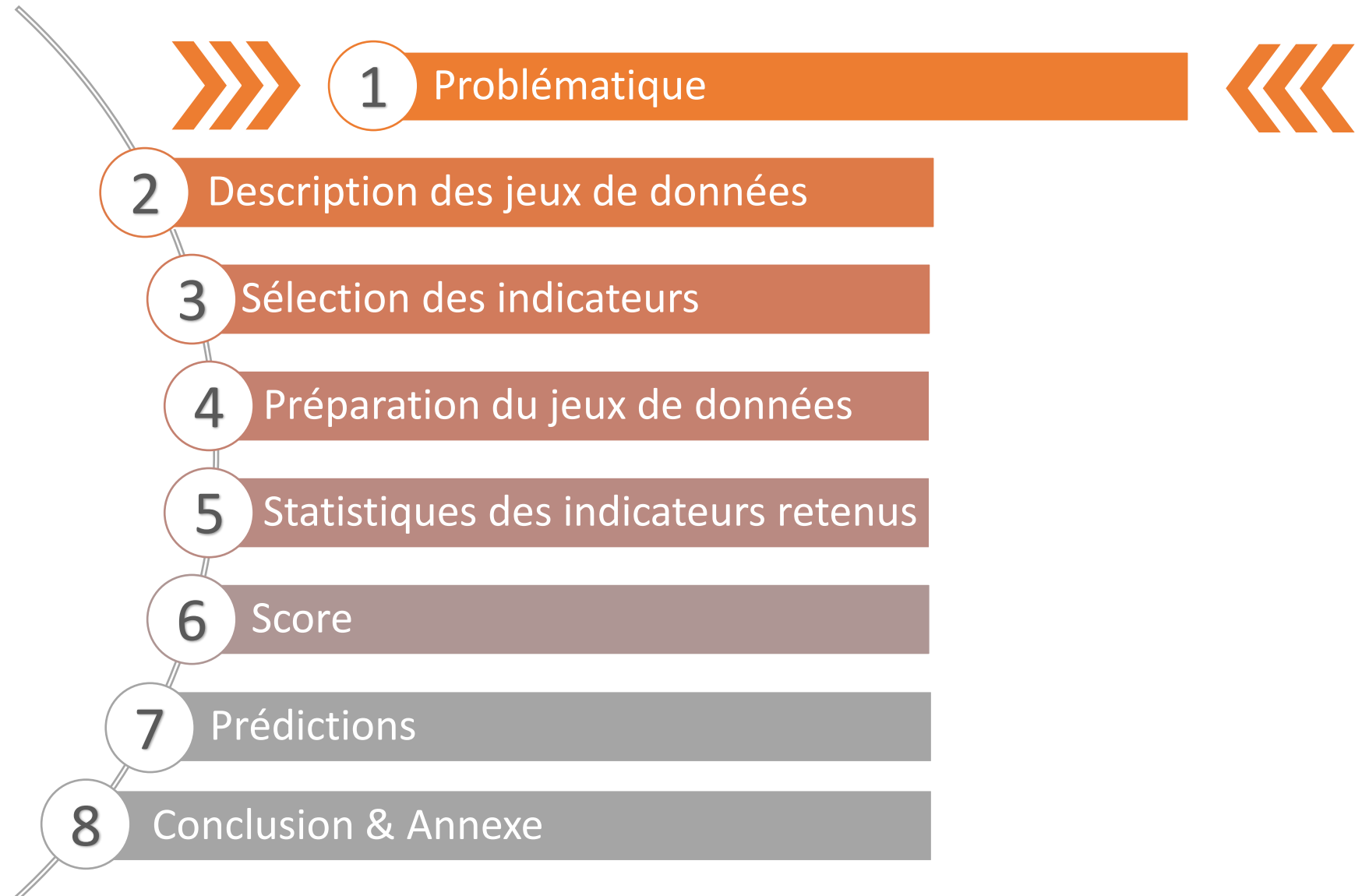
# Pré Analyse Exploratoire

Projet d'expansion à l'international  academy

Source :  **THE WORLD BANK**

<https://datacatalog.worldbank.org/dataset/education-statistics>





Projet d'expansion à l'international de formation en ligne de niveau lycée et supérieur

## Pour lever les interrogations ...

- quels sont les **pays** avec un fort potentiel de clients pour nos services ?
- Pour chacun de ces pays, quelle sera **l'évolution** de ce potentiel de clients ?
- Dans quels pays l'entreprise doit-elle opérer en **priorité** ?

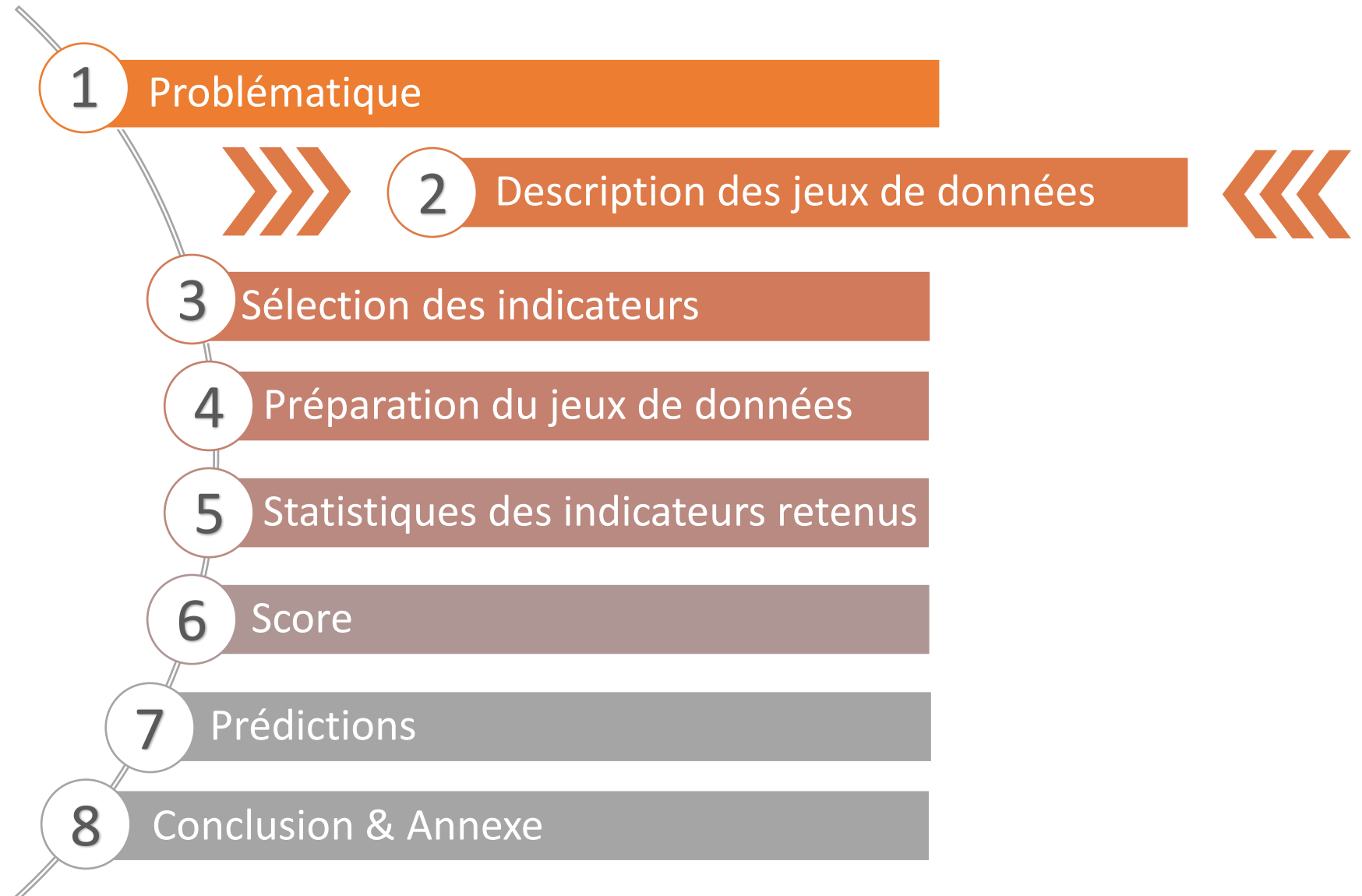


THE WORLD BANK

... pré-analyser les données de la banque mondiale du portail EdStat



- Informations pertinentes? De qualité? Cohérentes?
- Indicateurs statistiques pour départager les pays?
- Conclusion : pays à fort potentiel ? Evolution? Priorisation?



# Description

Objectif : Qualité, Cohérentes? : 



academy



WORLD BANK



EdStatsData.csv

Evolution des indicateurs par pays ou regroupement sur plusieurs années avec projection

886930 lignes, 70 variables (entre 73% et 100% NaN, , doublon)



EdStatsCountry.csv

Informations géographiques sur les pays par régions, données économiques globales et dates de référence

241 lignes, 32 variables (~30% NaN, doublon)



EdStatsSeries.csv

Informations socio-éduco-économiques des indicateurs par thème

3665 lignes, 21 variables (5 principales complètes)

EdStatsFootNote.csv

Informations sur l'année d'origine/incertitude des indicateurs par pays

241 lignes, 32 variables (principales complètes)

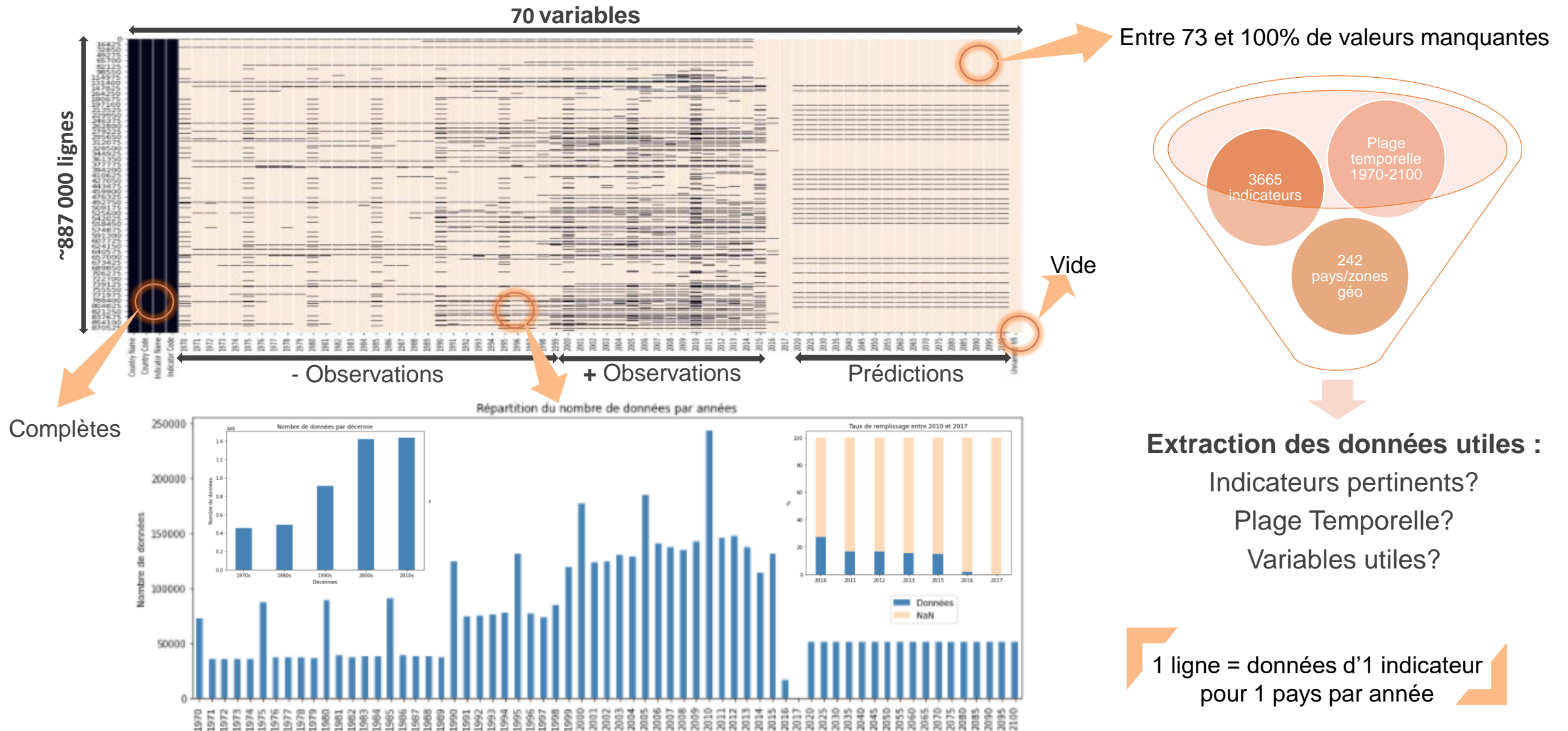
EdStatsCountry-Series.csv

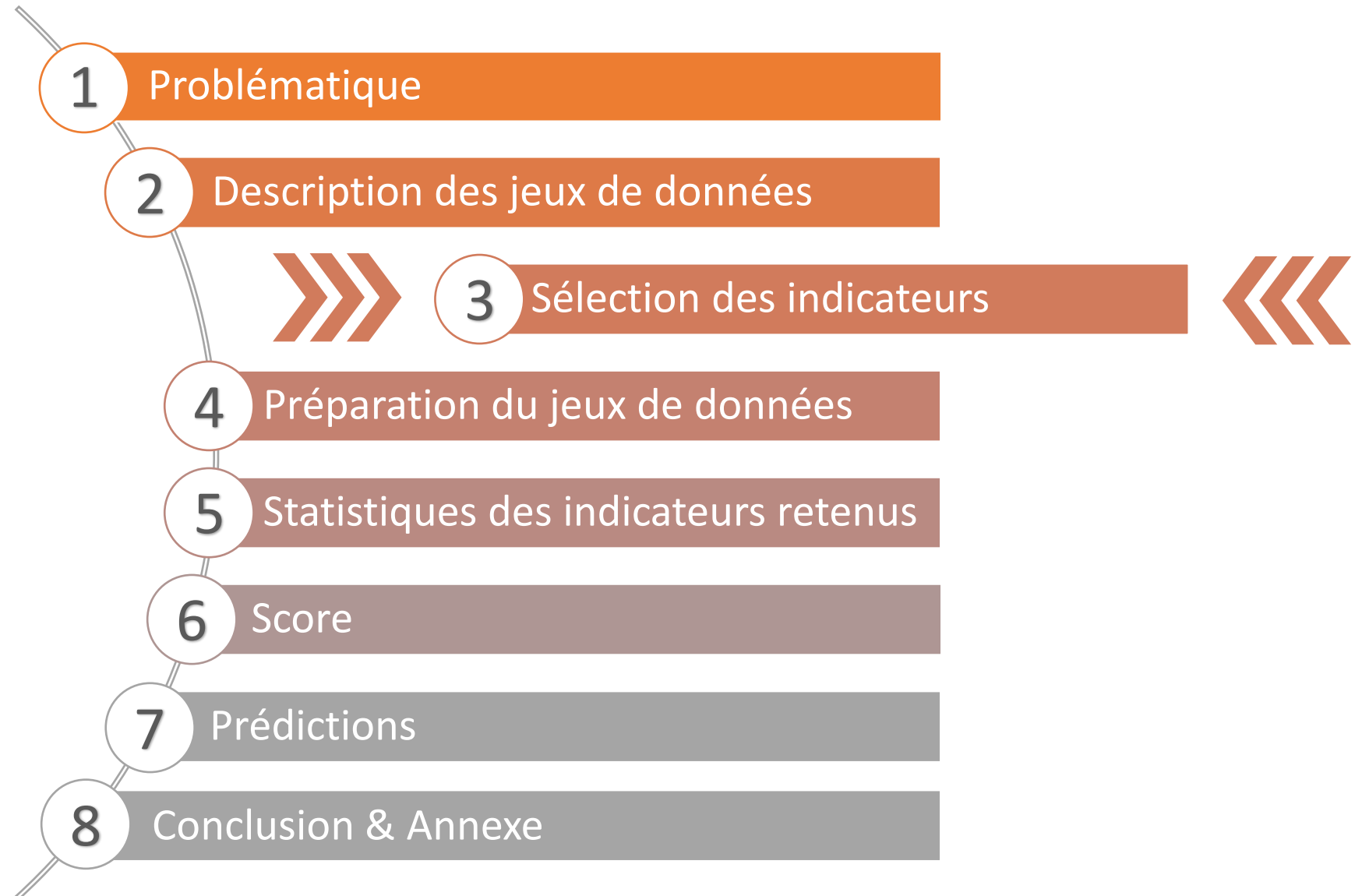
Informations sur la source des indicateurs par pays

613 lignes, 4 variables (principales complète)

Données tracées, sources, années de mise à jour, incertitudes

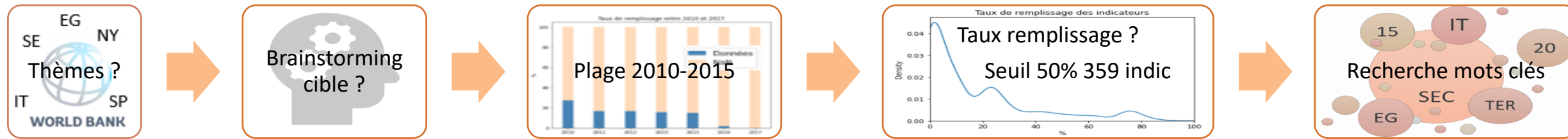
# Le jeu de données EdStatsData.csv







## Stratégie : 3665 indicateurs à analyser



### Démographique

- Population des 15-19 ans : lycée
- Population des 20-24 ans : supérieur

### Indicateurs retenus

**SP.POP.1524.TO.UN**

### Economique

- Revenus par habitant, niveau de vie
- Comparaison devises, inflation..

**NY.GNP.PCAP.PP.CD**

### Educatif

- 15-19 et 20-24 ans, réussite
- Taux de scolarisation secondaire, tertiaire

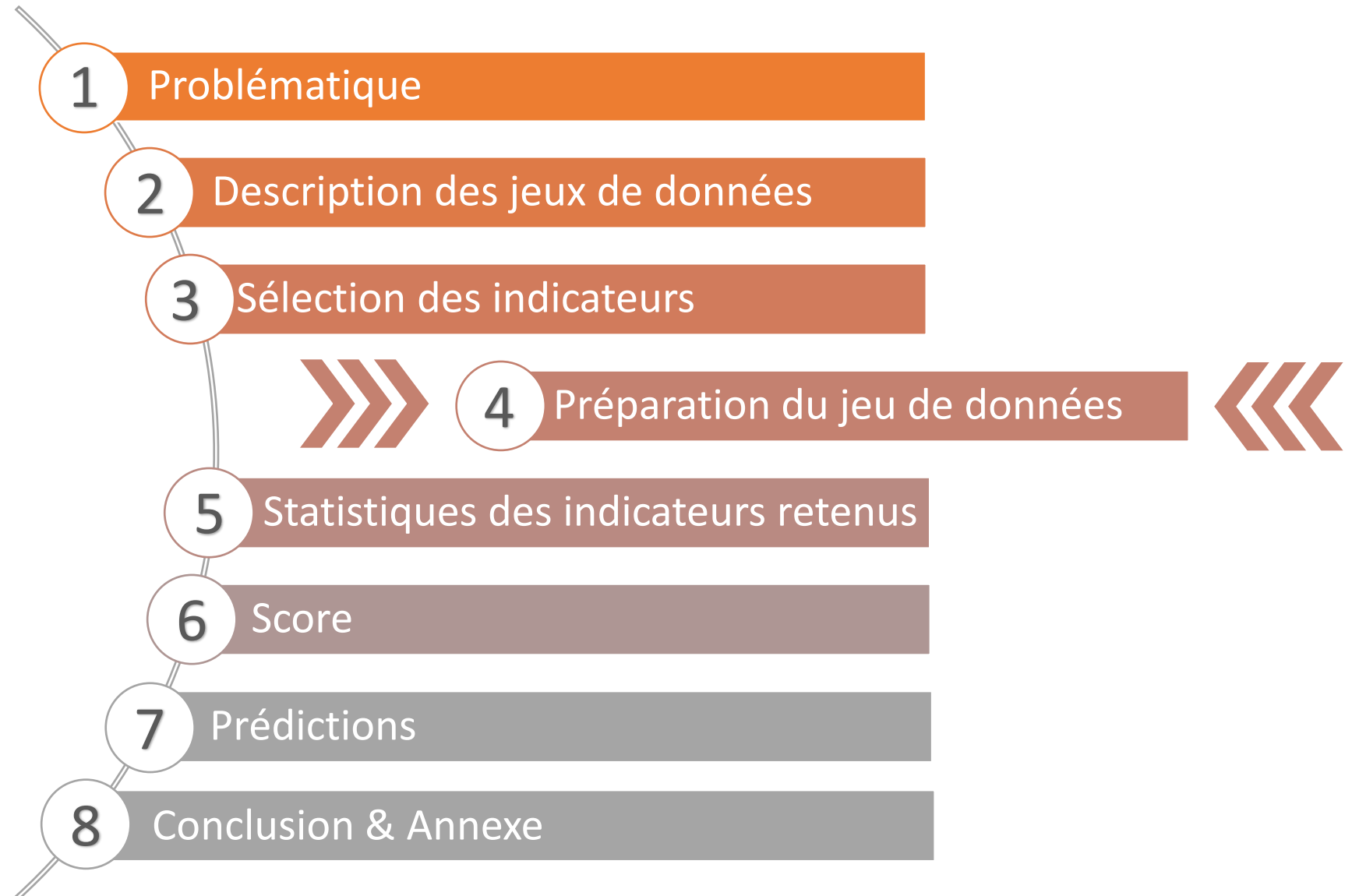
**SE.SEC.ENRR**

**SE.TER.ENRR**

### Numérique

- Accès à internet
- Accès à l'électricité, PC personnel

**IT.NET.USER.P2**



# Préparation

Objectif : Pertinentes? : ☒



## EdStatsData.csv

## EdStatsData.csv

## Dataset d'analyse statistique des indicateurs 2015

1  
Réduction

886930\*63 - 242

Suppression de la variable 'Unnamed: 69' vide

886930\*10 - 242

Réduction temporelle : 2010-2015

1452\*10 - 242

Filtre par indicateurs retenus

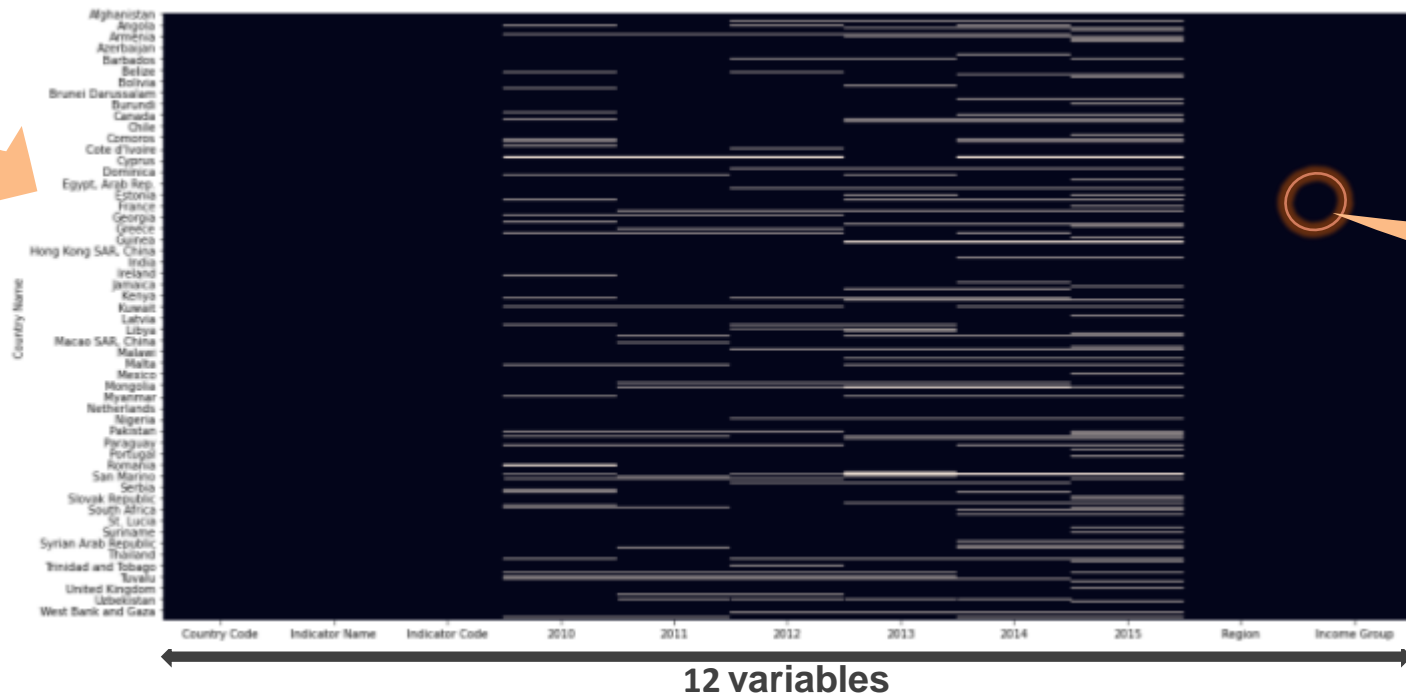
1302\*10 - 217

Suppression des codes pays ISO hors norme

916\*10 - 210

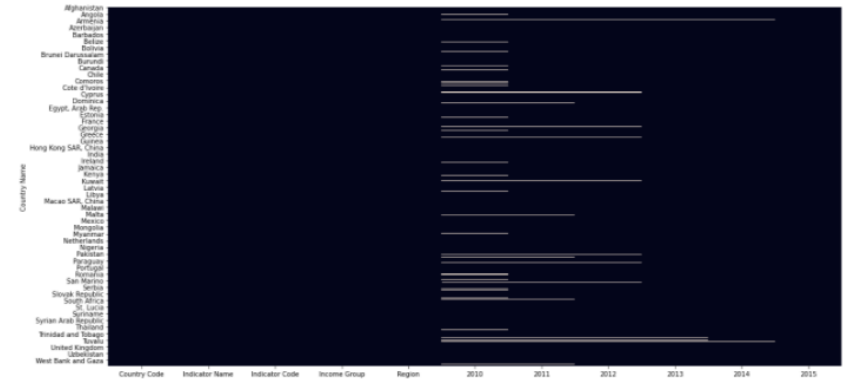
Suppression des indicateurs des pays avec aucune donnée pour toutes les années

210 pays



4

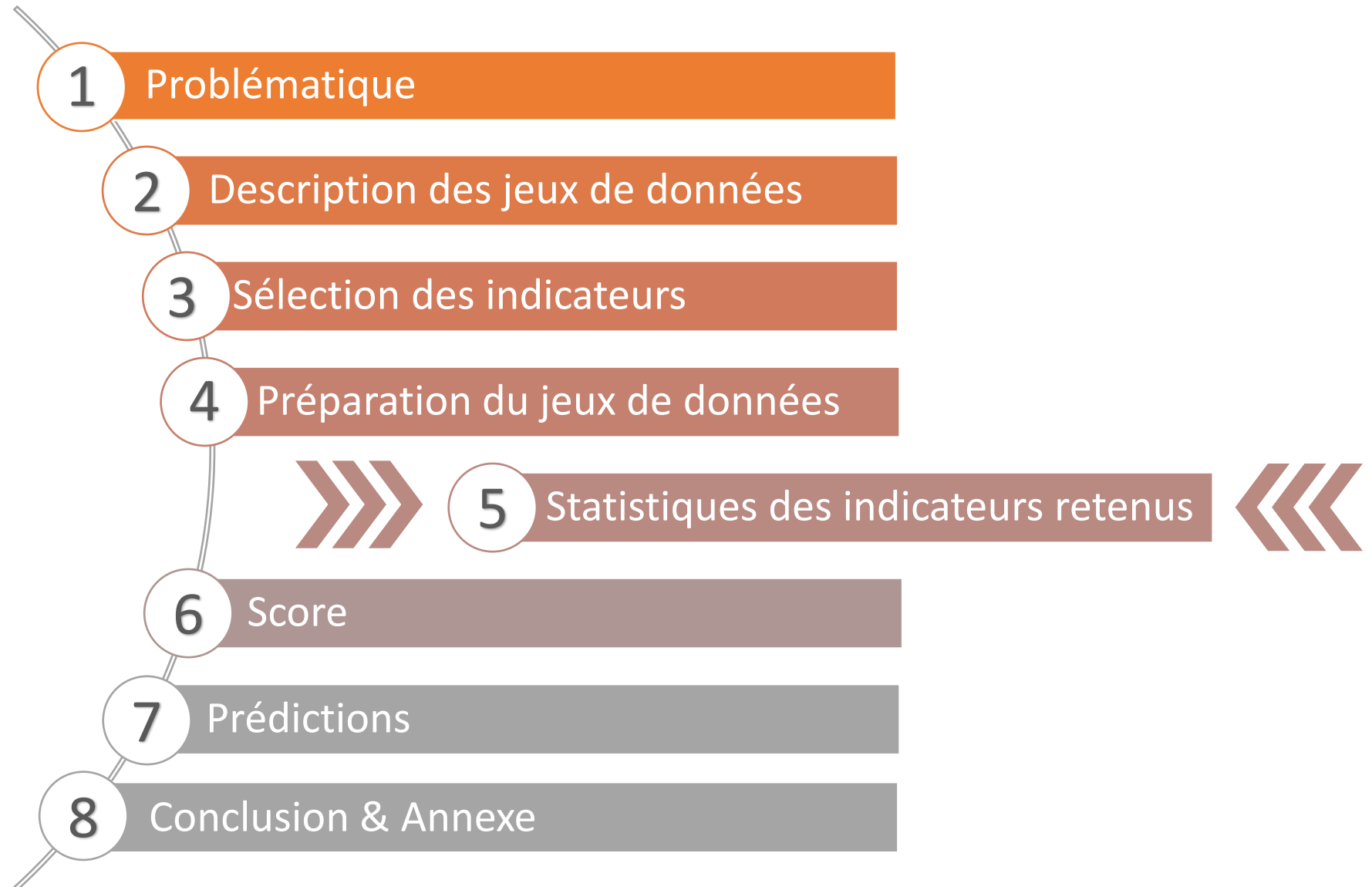
Réduction :  
5 indicateurs  
140 pays



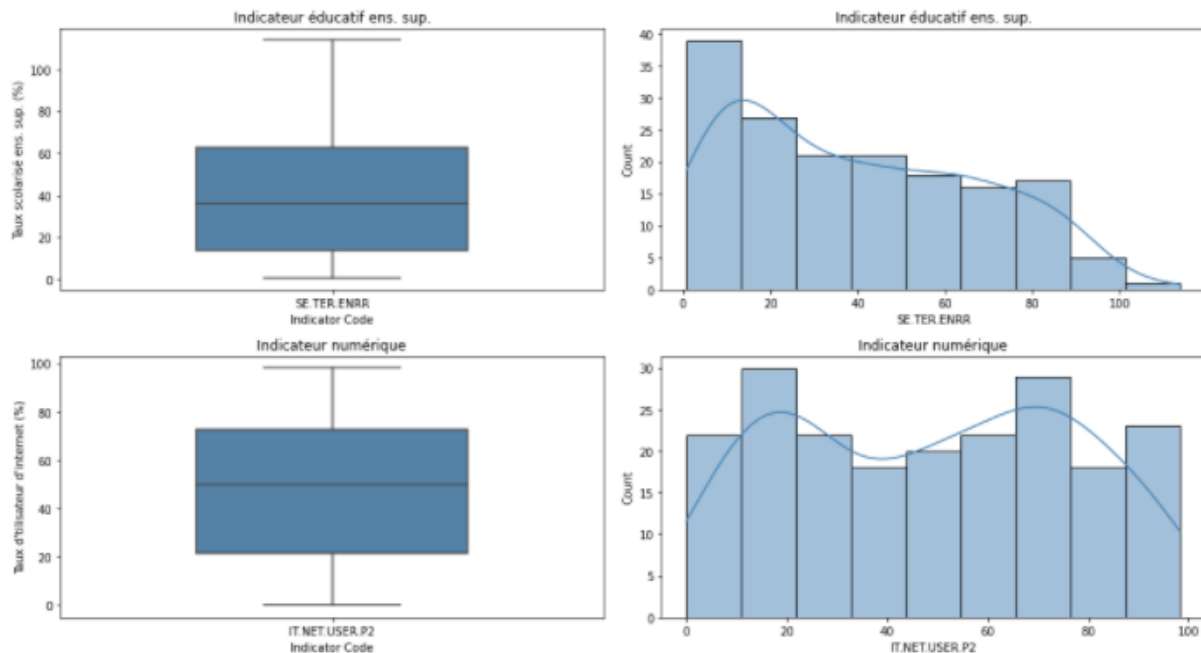
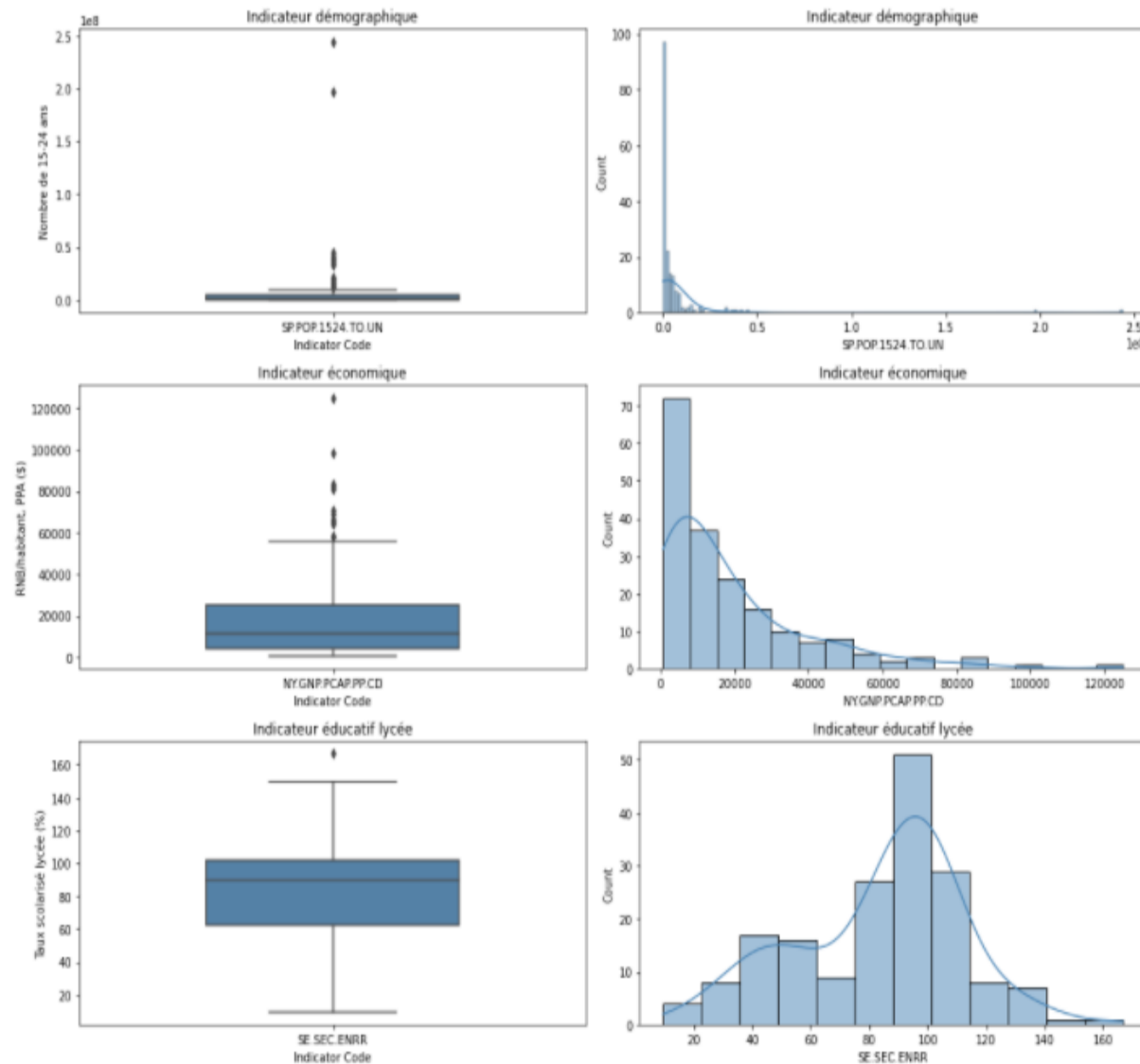
3 Imputation

2

Ajout des variables :  
région  
groupe de revenus

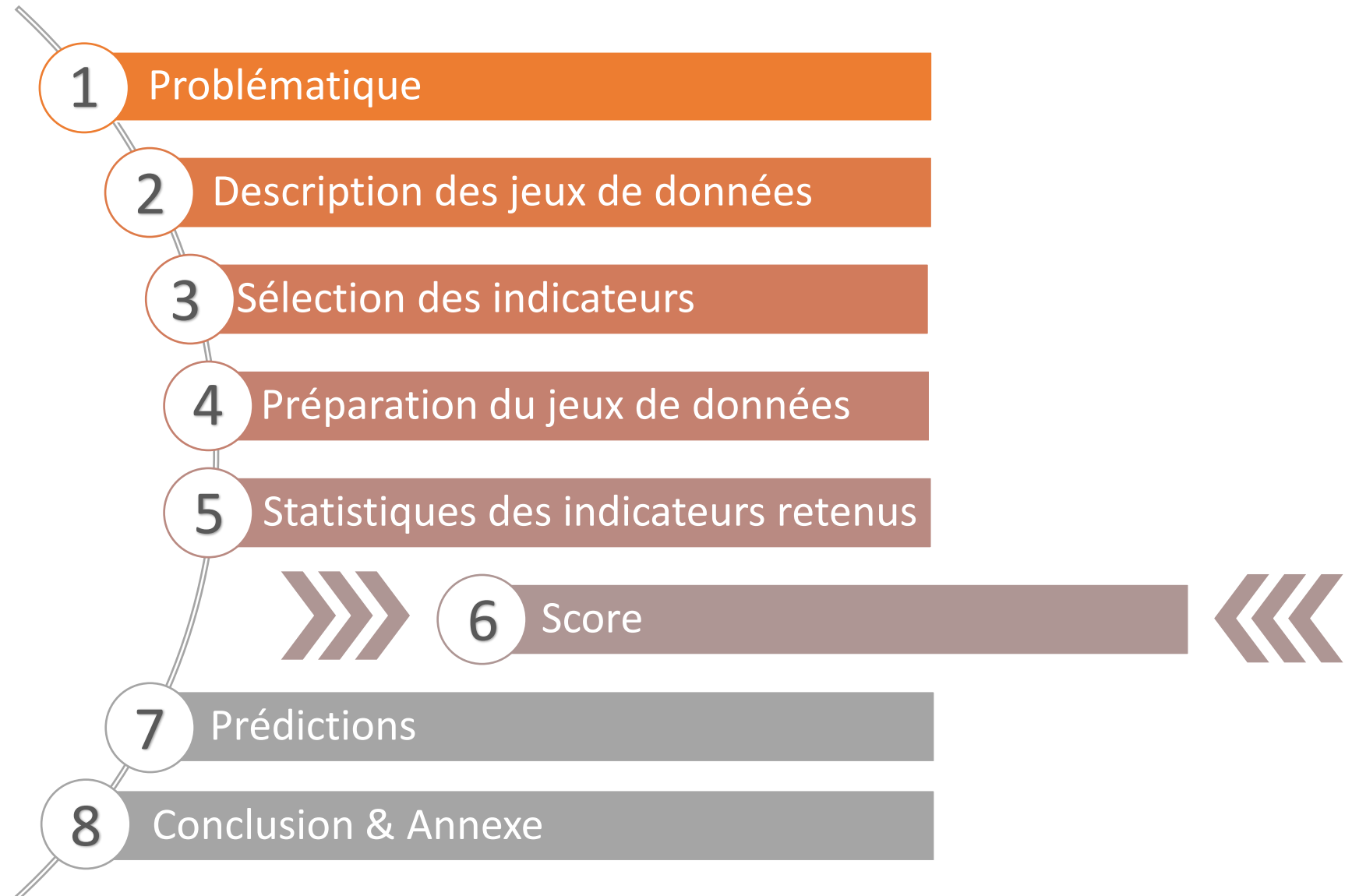


# Statistiques des indicateurs retenus



Desc	Stat_num	Stat_eco	Stat_edu_sec	Stat_edu_ter	Stat_pop
mean	48.63017	1.90438e+04	83.82360	39.49392	6.65499e+06
median	50.21966	1.16450e+04	89.88169	36.22854	1.33104e+06
var	811.61123	4.28604e+08	815.06711	775.82922	5.74750e+14
std	28.48879	2.07028e+04	28.54938	27.85371	2.39739e+07
skew	-0.00535	1.96926e+00	-0.35019	0.41960	8.28032e+00
kurtosis	-1.26904	4.77319e+00	-0.10257	-1.00113	7.43375e+01

Unités différentes, grandeurs différentes, Comparer ?  
Mise à l'échelle



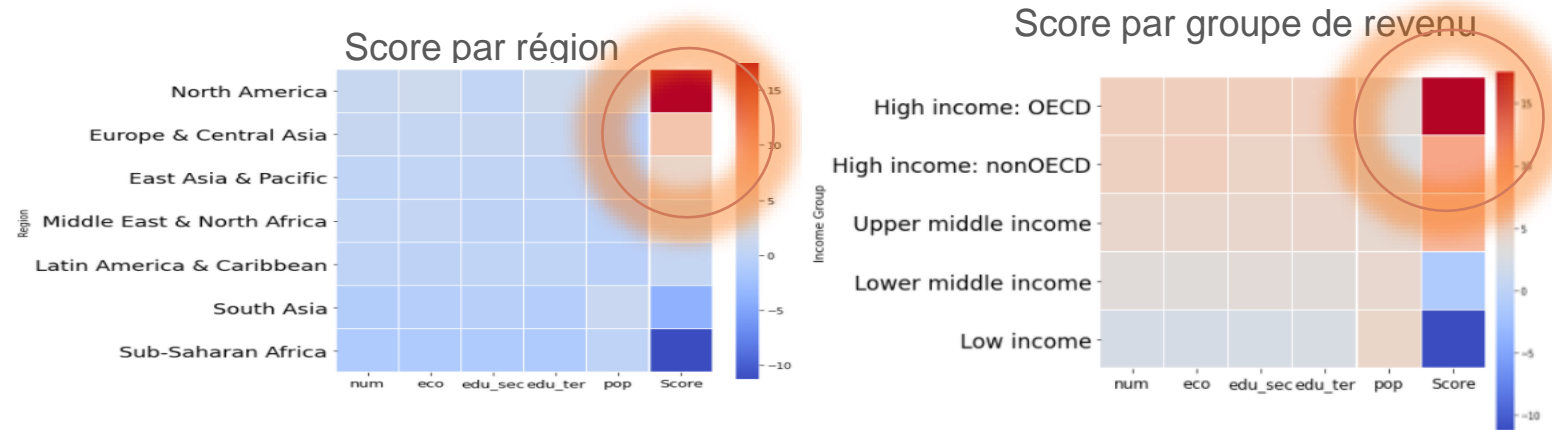
# Score

Objectif : Comparer, fort potentiel? : ☒

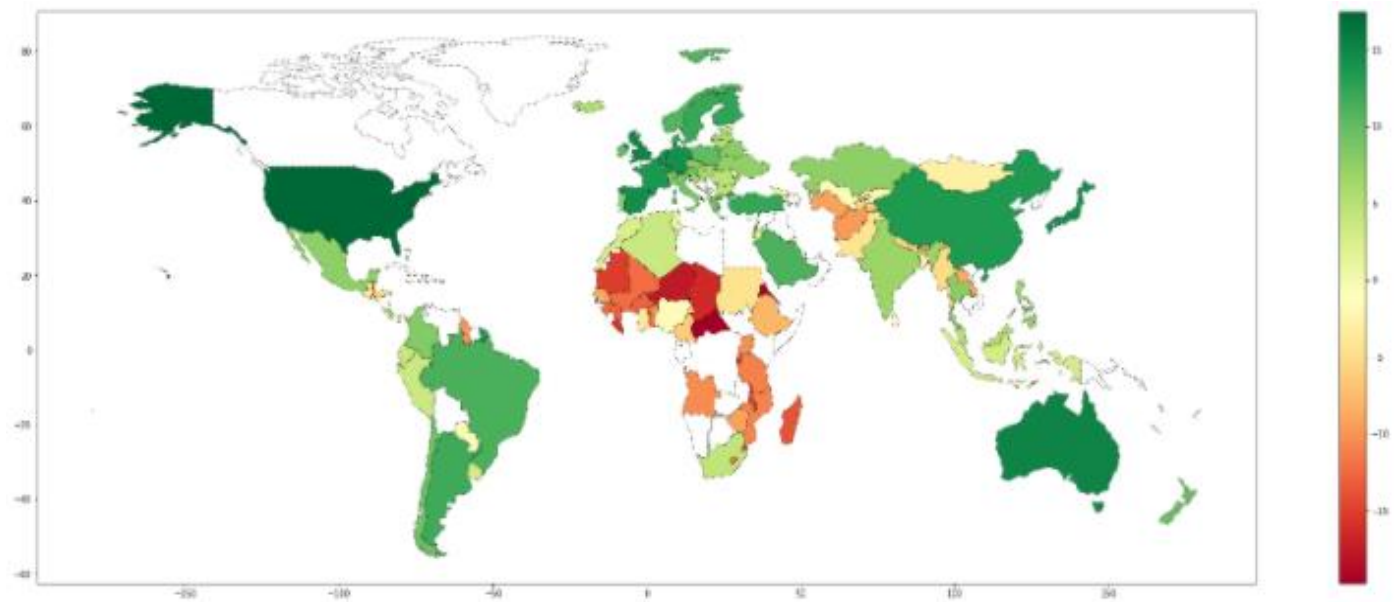
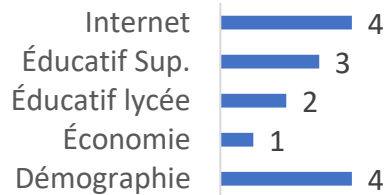


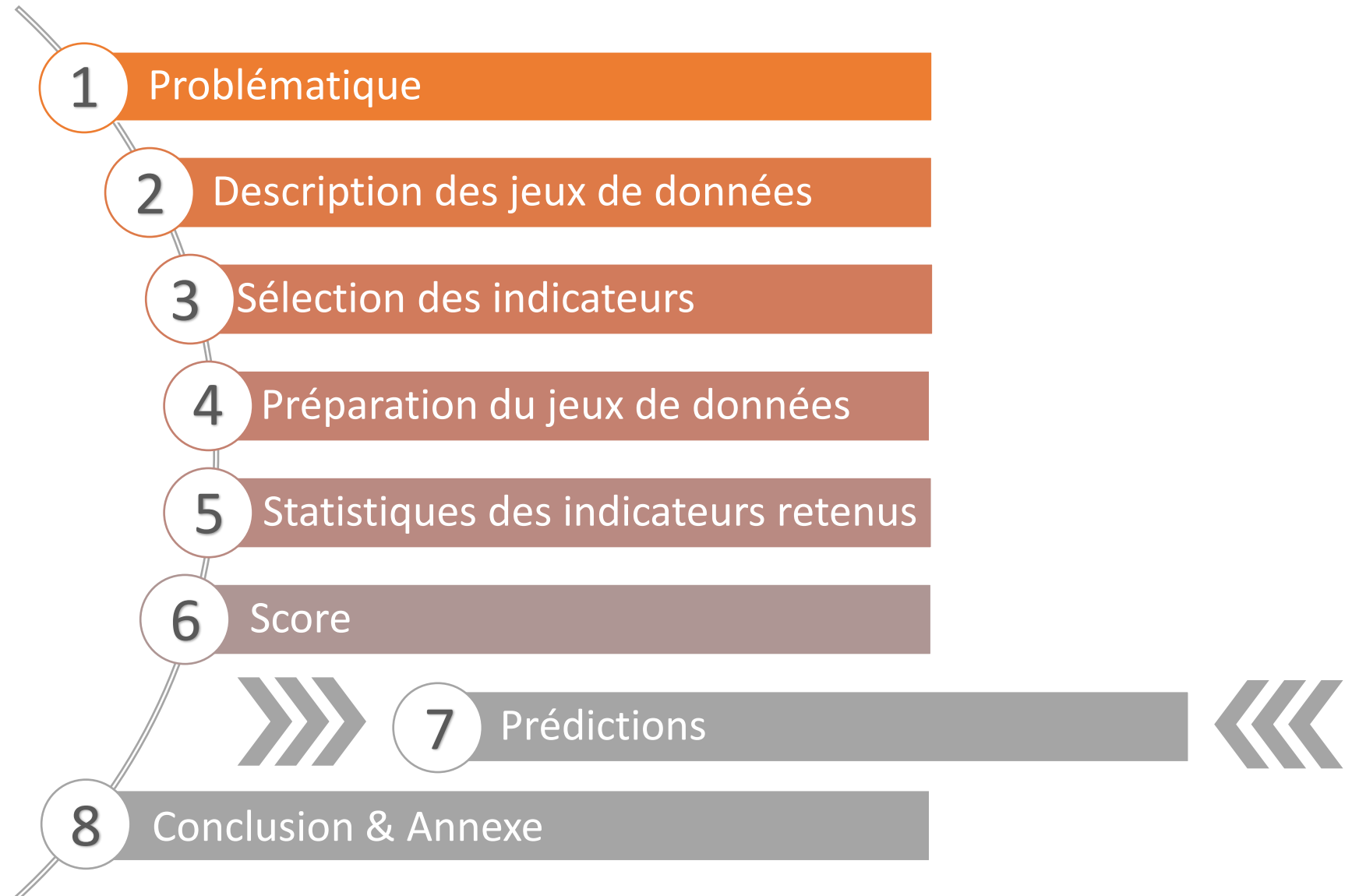
De 140 pays candidats...

... au top 10 des pays attractifs



	Place Provisoire	Pays	Score
	1	United States	17.49887
	2	Australia	15.17283
	3	United Kingdom	15.16156
	4	Japan	14.78471
	5	Korea, Rep.	14.67791
	6	Belgium	14.54049
	7	Spain	14.53744
	8	Netherlands	14.33014
	9	Germany	14.18948
	10	Russian Federation	14.02221







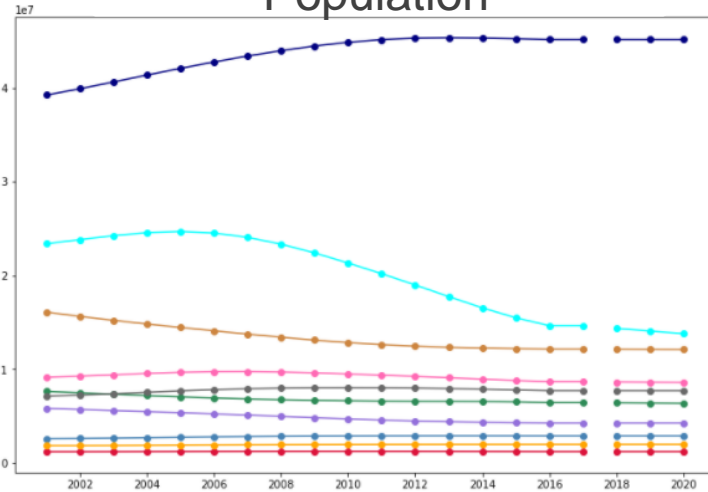
# Prédictions

Objectif : Evolution? : ☒

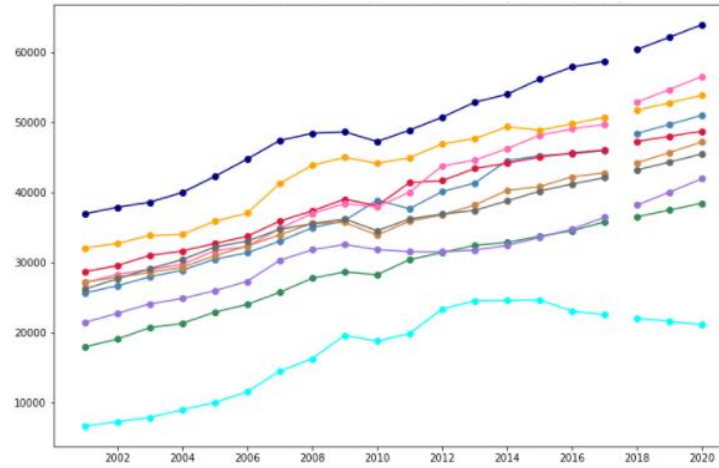


academy

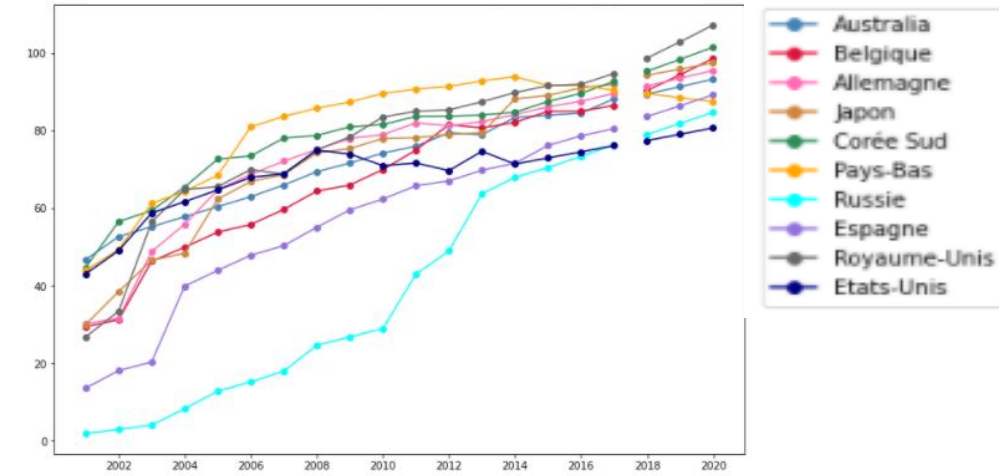
## Population



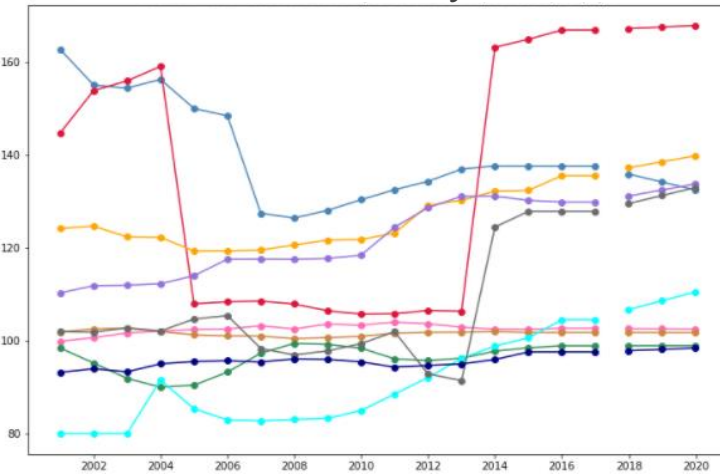
## Économie



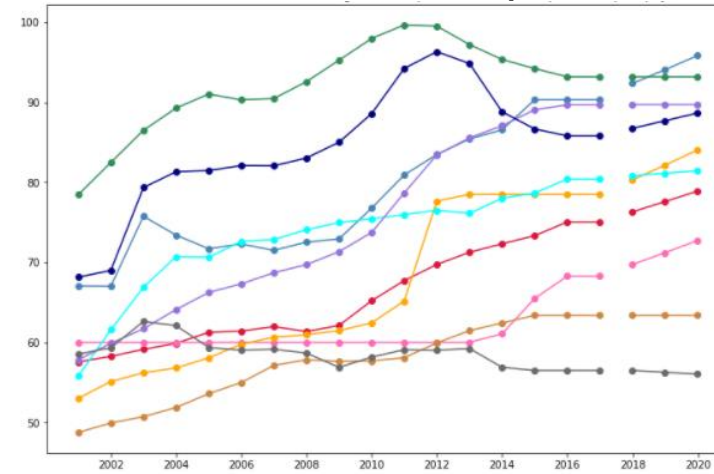
## Internet



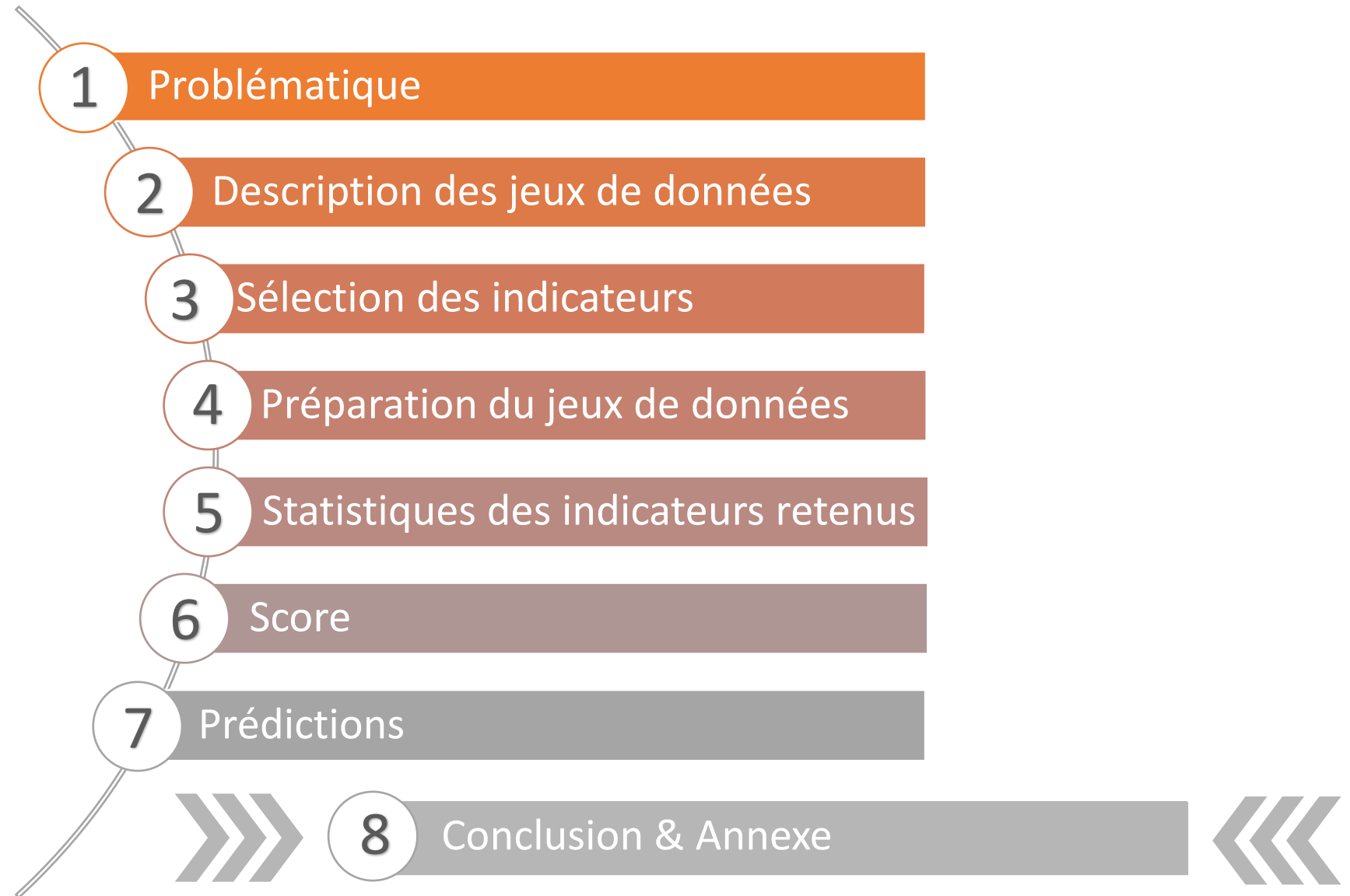
## Éducation lycée



## Éducation ens. supérieur

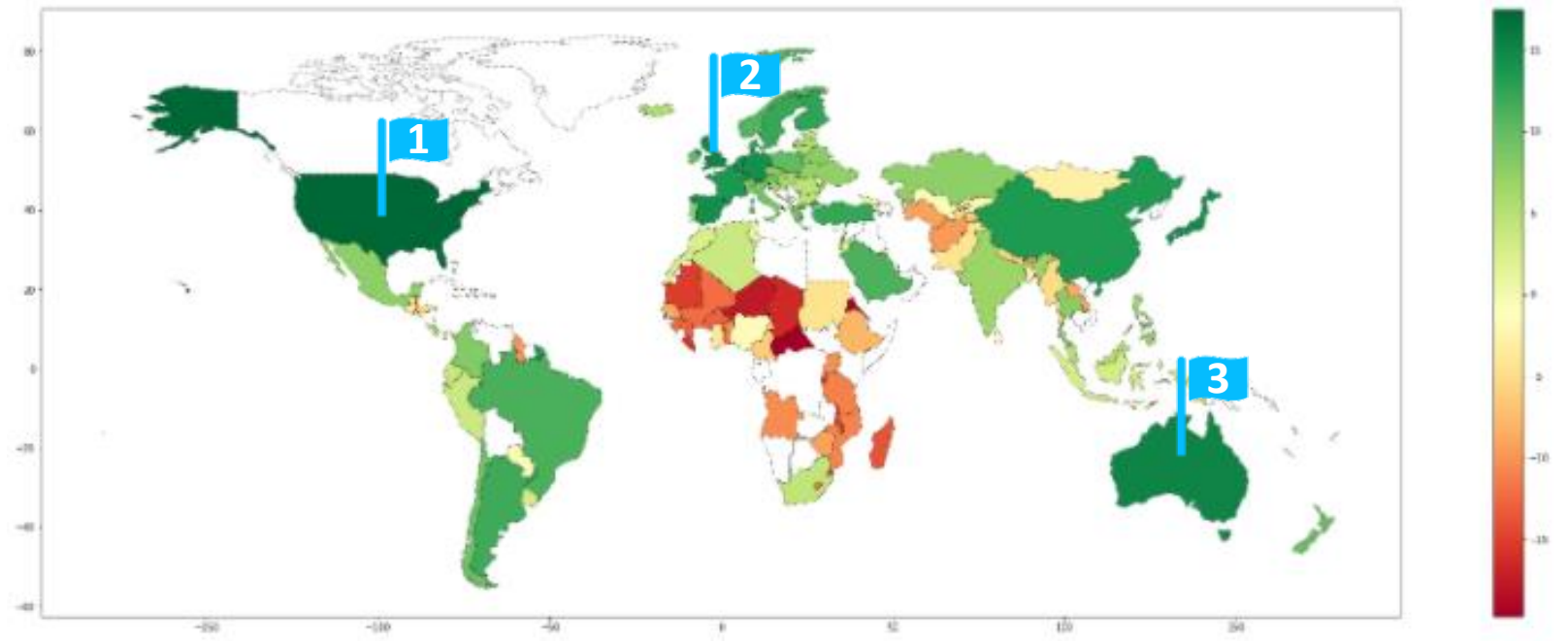
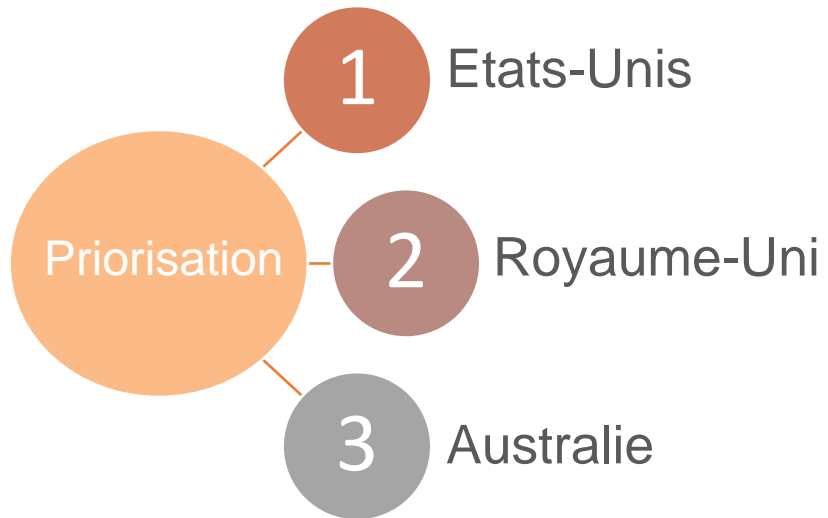


	Score	Place Provisoire	Score Prédit	Place Prévisionnelle
Pays				
United States	17.49887	1	2.14031	4
United Kingdom	15.16156	3	4.02281	2
Australia	15.17283	2	2.79532	3
Korea, Rep.	14.67791	5	4.91644	1
Japan	14.78471	4	-0.66913	5
Spain	14.53744	7	-0.99597	7
Germany	14.18948	9	-0.79066	6
Belgium	14.54049	6	-2.78101	8
Netherlands	14.33014	8	-3.72200	9
Russian Federation	14.02221	10	-4.91610	10



# Conclusion

Objectif : Priorisation? : ☒



Jeu de données plus récent, plus de données éducatives

Vérifier l'hypothèse de travail avec les décideurs

Stratégie d'entreprise : langue, marché cible, proximité géographique d'implantation

Autres indicateurs, Concurrence

Forte impression des pays d'Asie de l'Est et du Pacifique

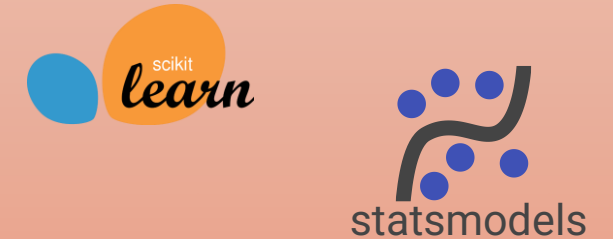
## Environnement



## Librairies de base



## API



## Outil de bureau



## Visualisation

