

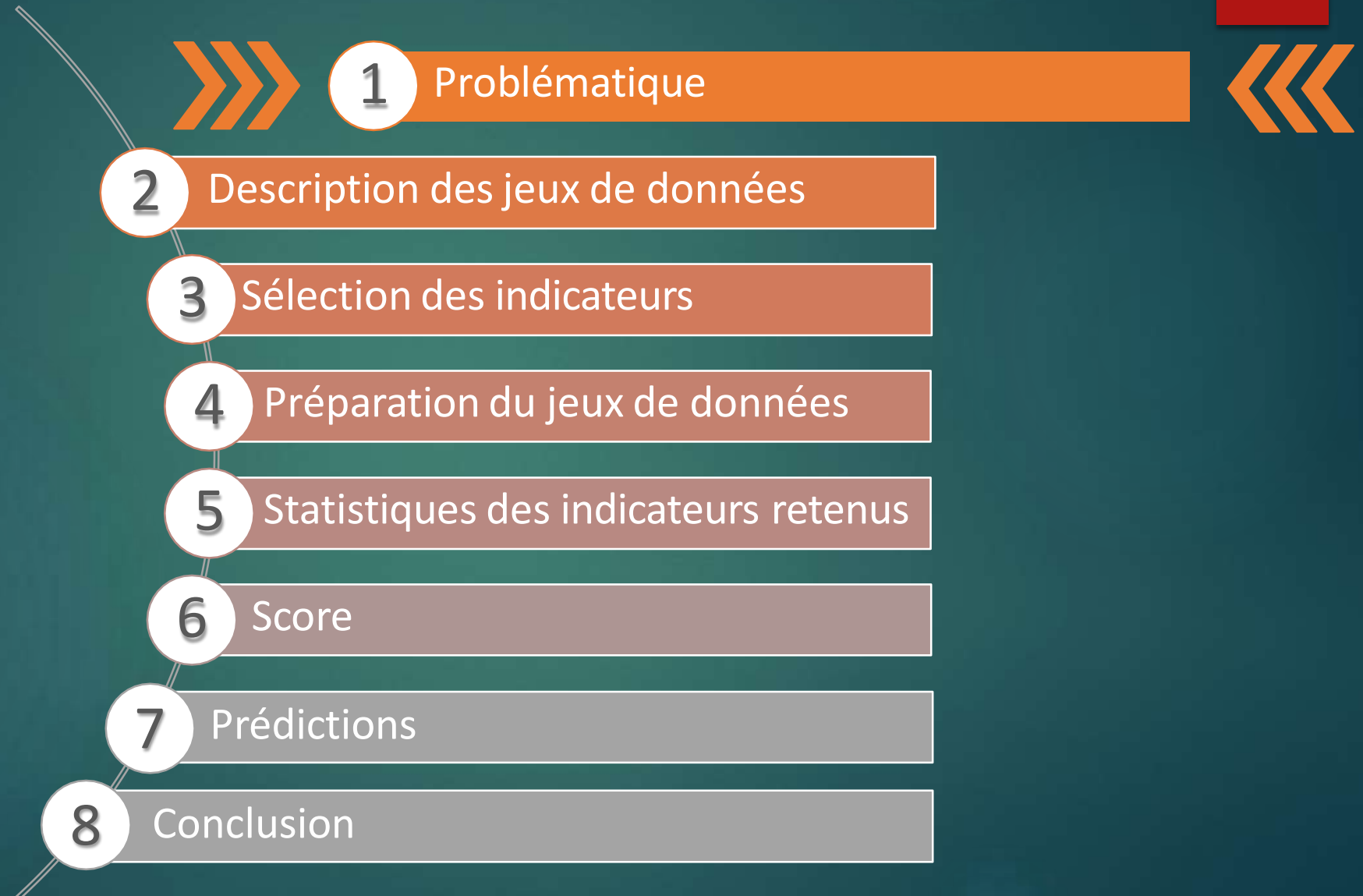


Projet 2 : Analyser des données de systèmes éducatifs

OPENCLASSROOMS

Présenté par : **Lynda HADJEMI**





- ❖ Academy : Start-up de la EdTech.
- ❖ Contenus de formation en ligne pour un public de niveau lycée et université.
- ❖ Projet d'expansion à l'international de formation en ligne de niveau lycée et supérieur.

Objectifs:

- Effectuer une analyse pré-exploratoire du jeu de données de la banque mondiale.
- Déterminer une liste de pays à fort potentiel pour une expansion à l'international.





Description



EdStatsData.csv

Evolution des indicateurs par pays ou regroupement sur plusieurs années avec projection(1970 à 2100) et 242 pays
886930 lignes, 70 variables (entre 73% et 100% NaN, , doublon)



EdStatsCountry.csv

Informations géographiques sur les pays par régions, quelques données manquantes , pas de doublons,
241 lignes, 32 variables



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 (643638 lignes, 5 variables (principales complètes))

EdStatsCountry-Series.csv

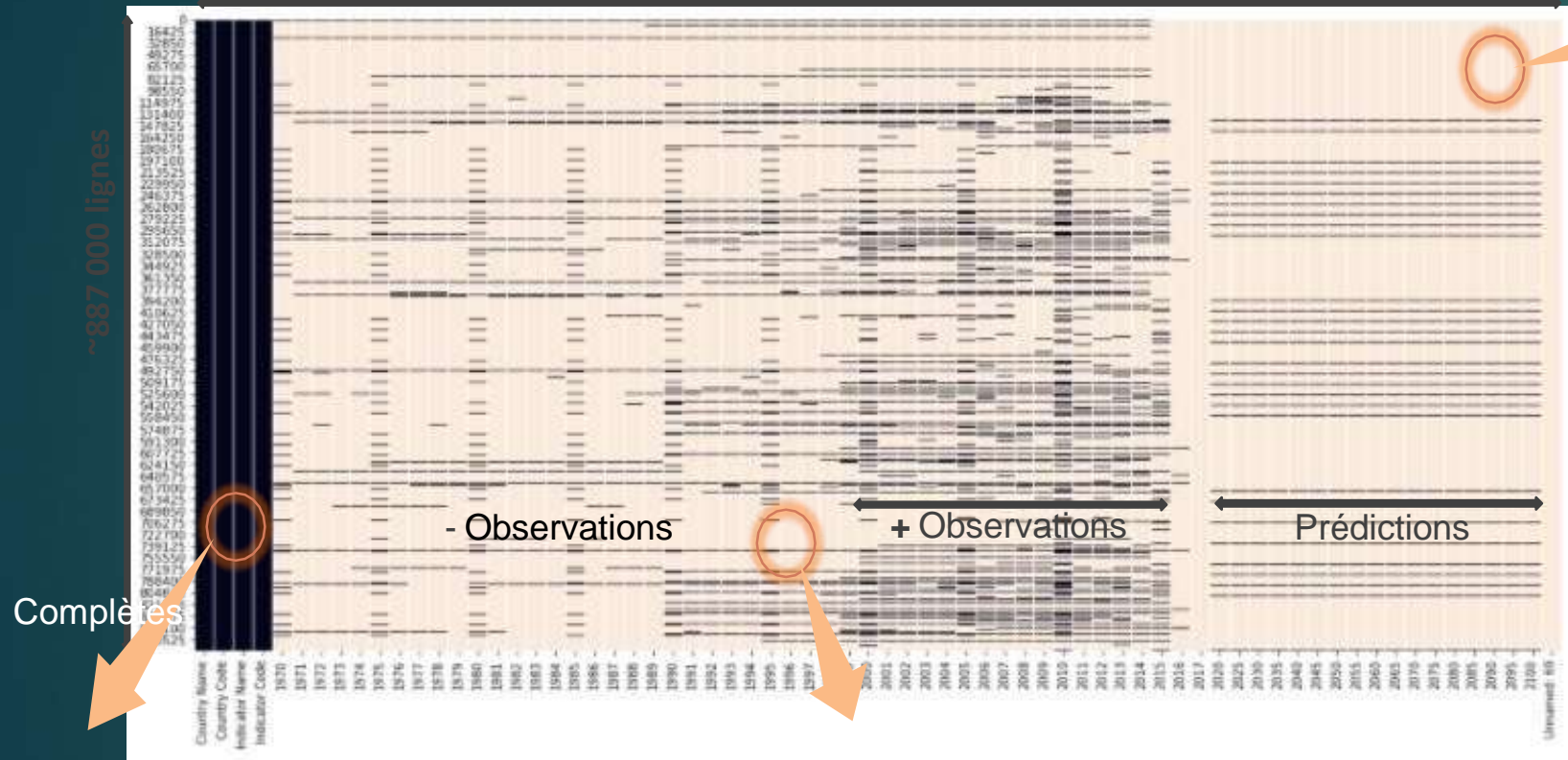
Informations sur la source des indicateurs par certains pays et certains indicateurs, 613 lignes, 4 variables (principales complète)



WORLD BANK

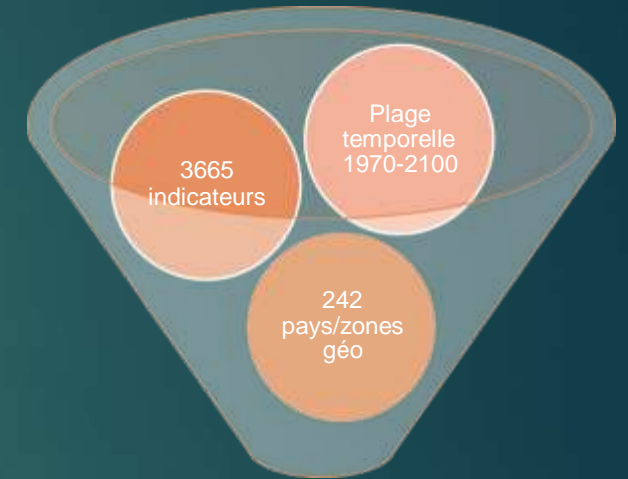
Le jeu de données EdStatsData.csv

70 variables



Entre 73 et 100% de valeurs manquantes

Vide

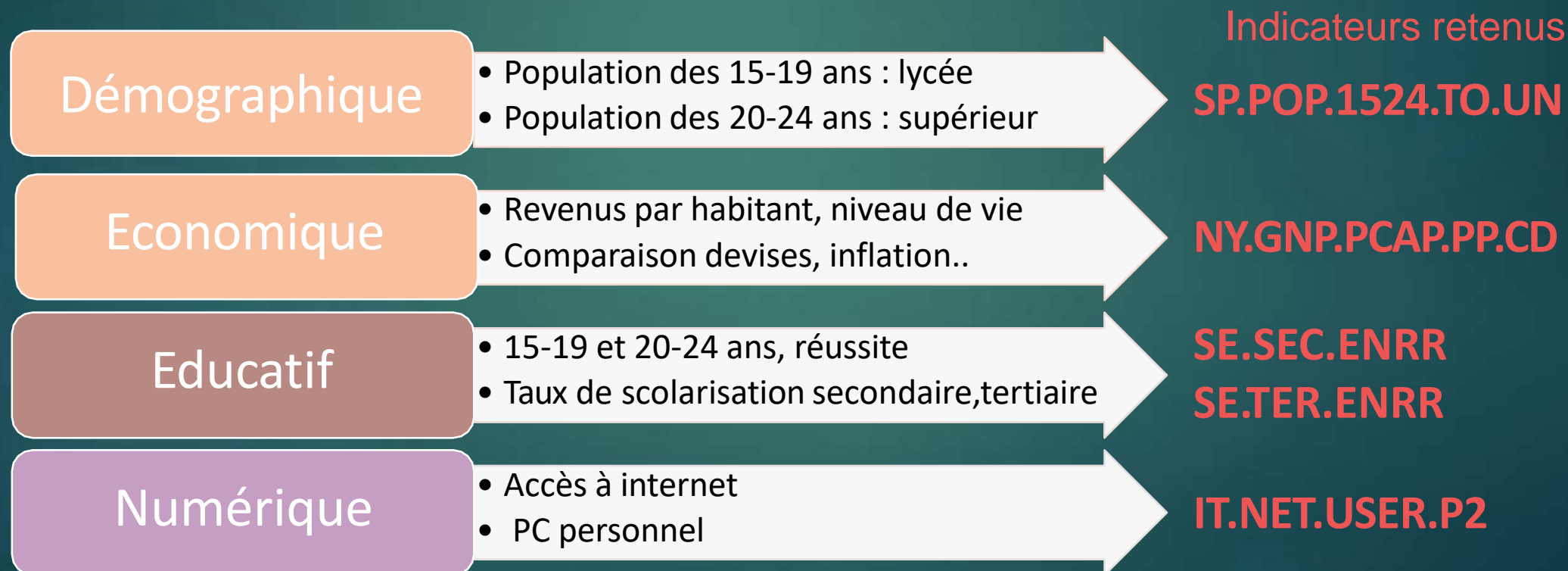


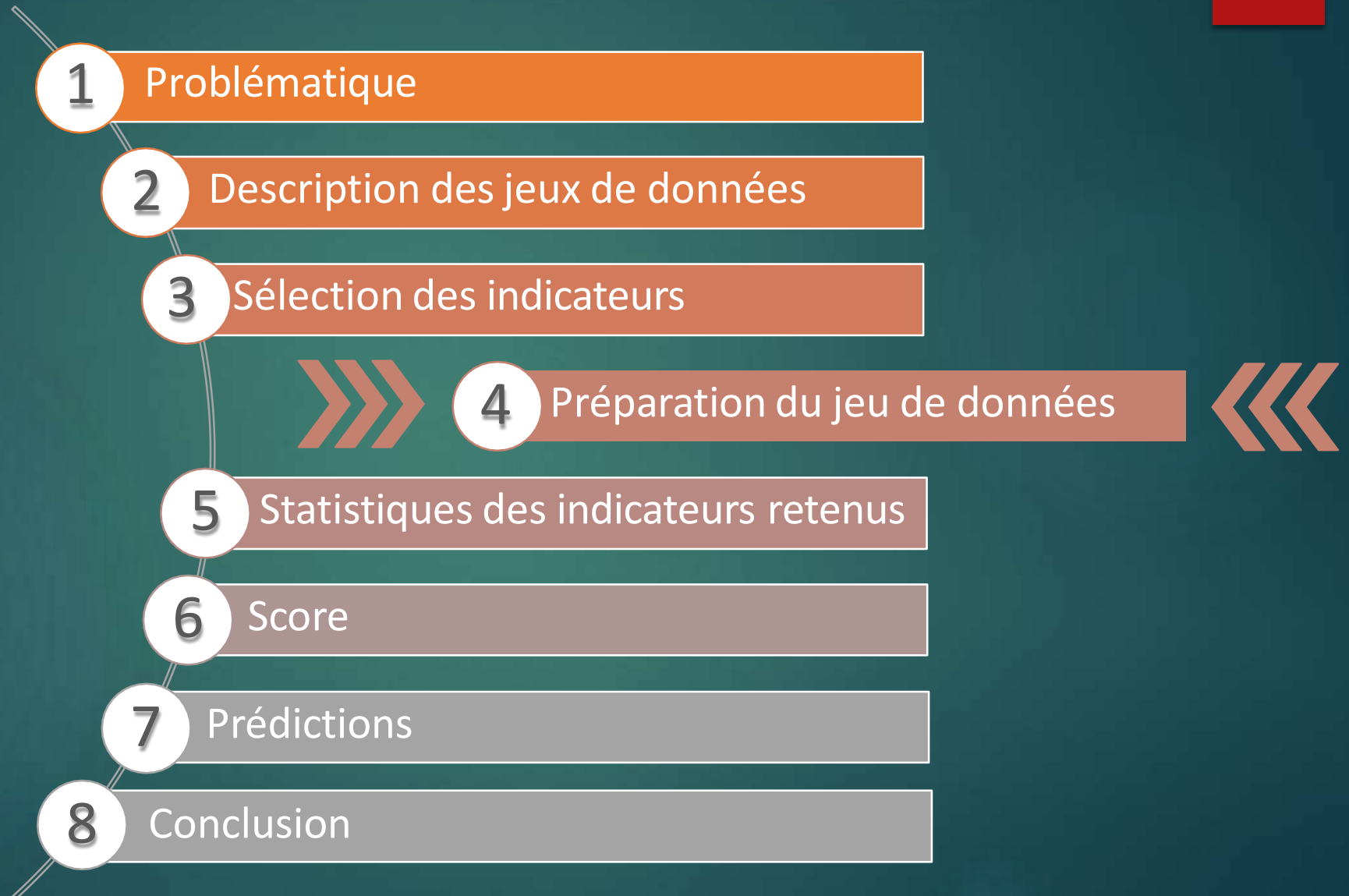
Extraction des données utiles :

Indicateurs pertinents?
Plage Temporelle?
Variables utiles?



Stratégie : 3665 indicateurs à analyser





EdStatsData.csv

EdStatsData.csv

Suppression de la variable 'Unnamed: 69' vide

Réduction : 2000-2015

Filtre par indicateurs retenus

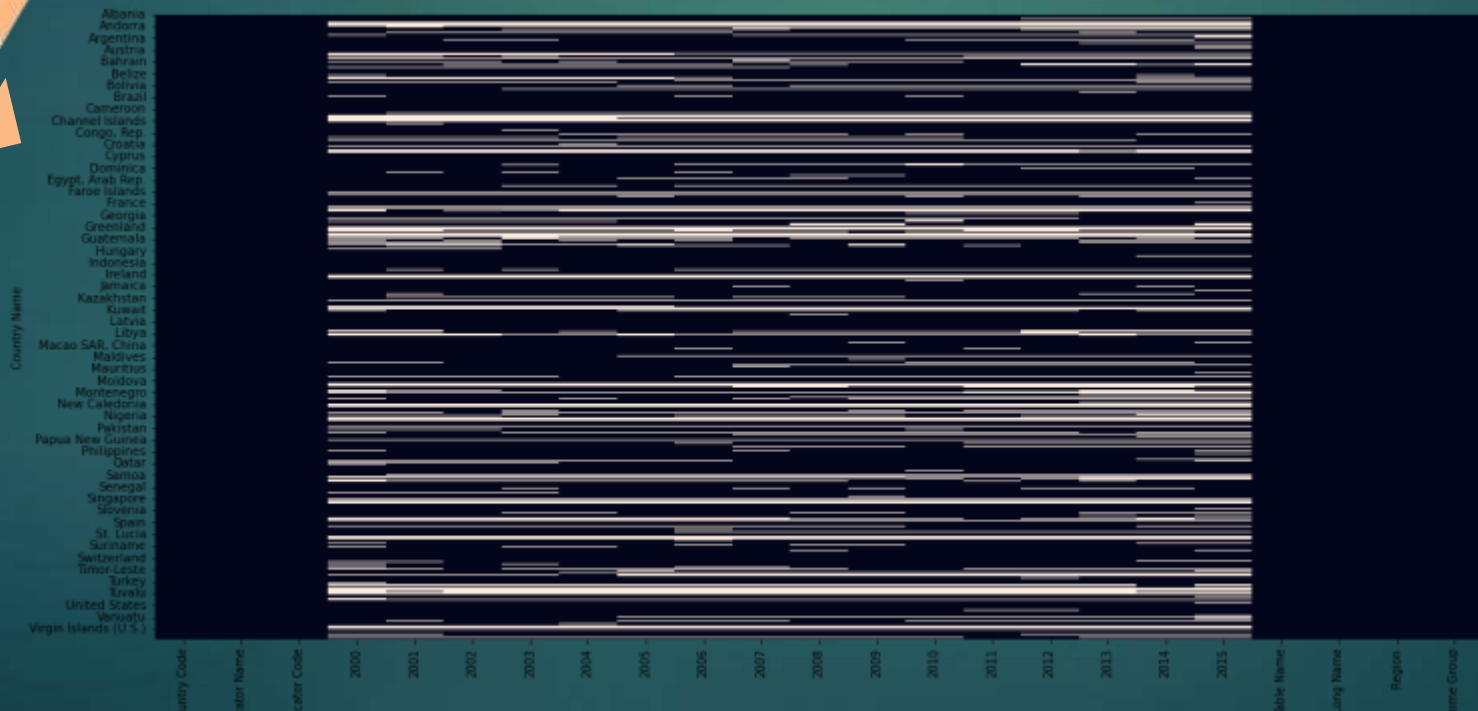
1210 ligne

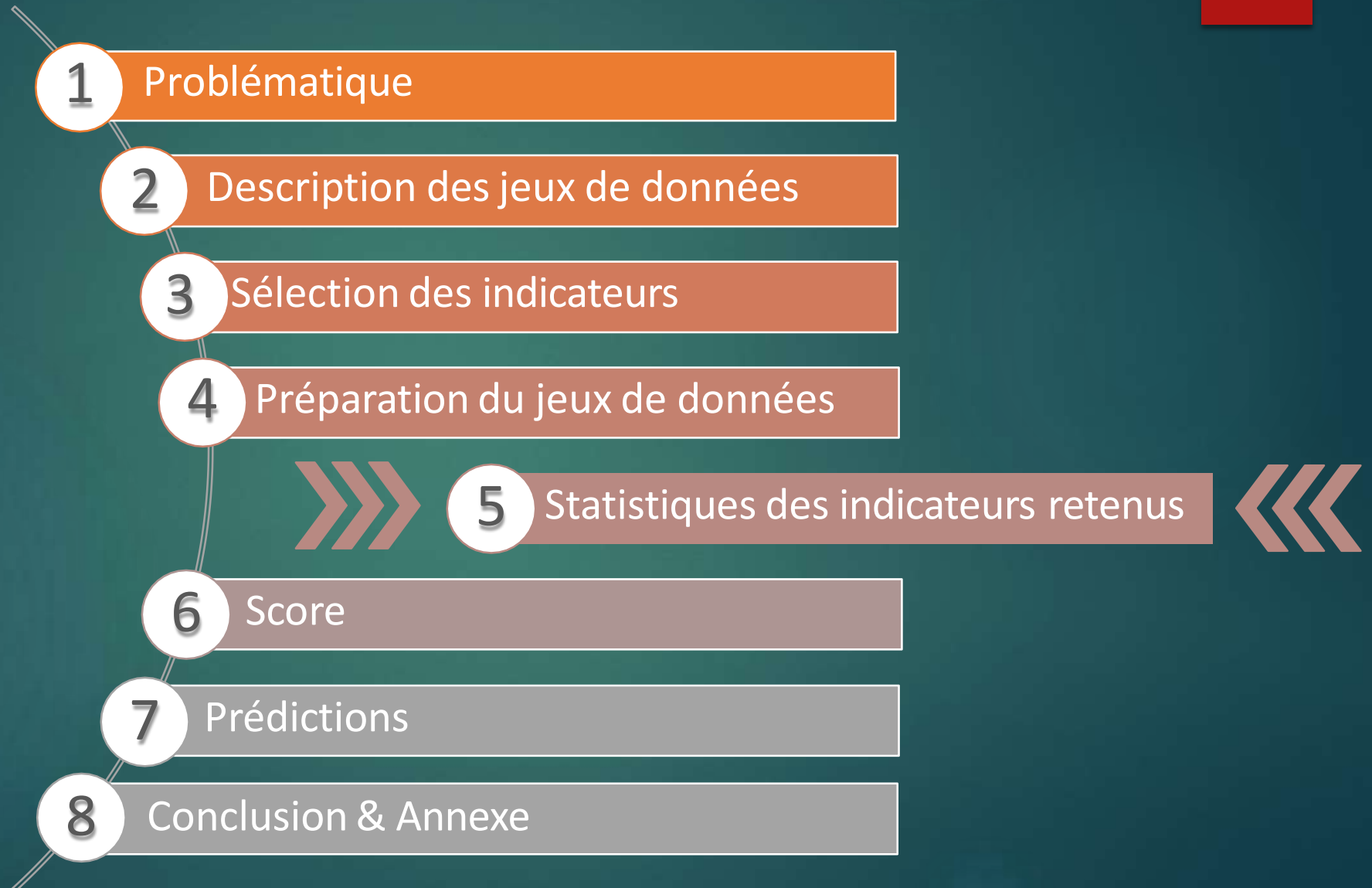
Filtre par revenus de pays

900 lignes

Suppression des indicateurs des pays avec aucune données ou manquante pour toutes les années

131 lignes



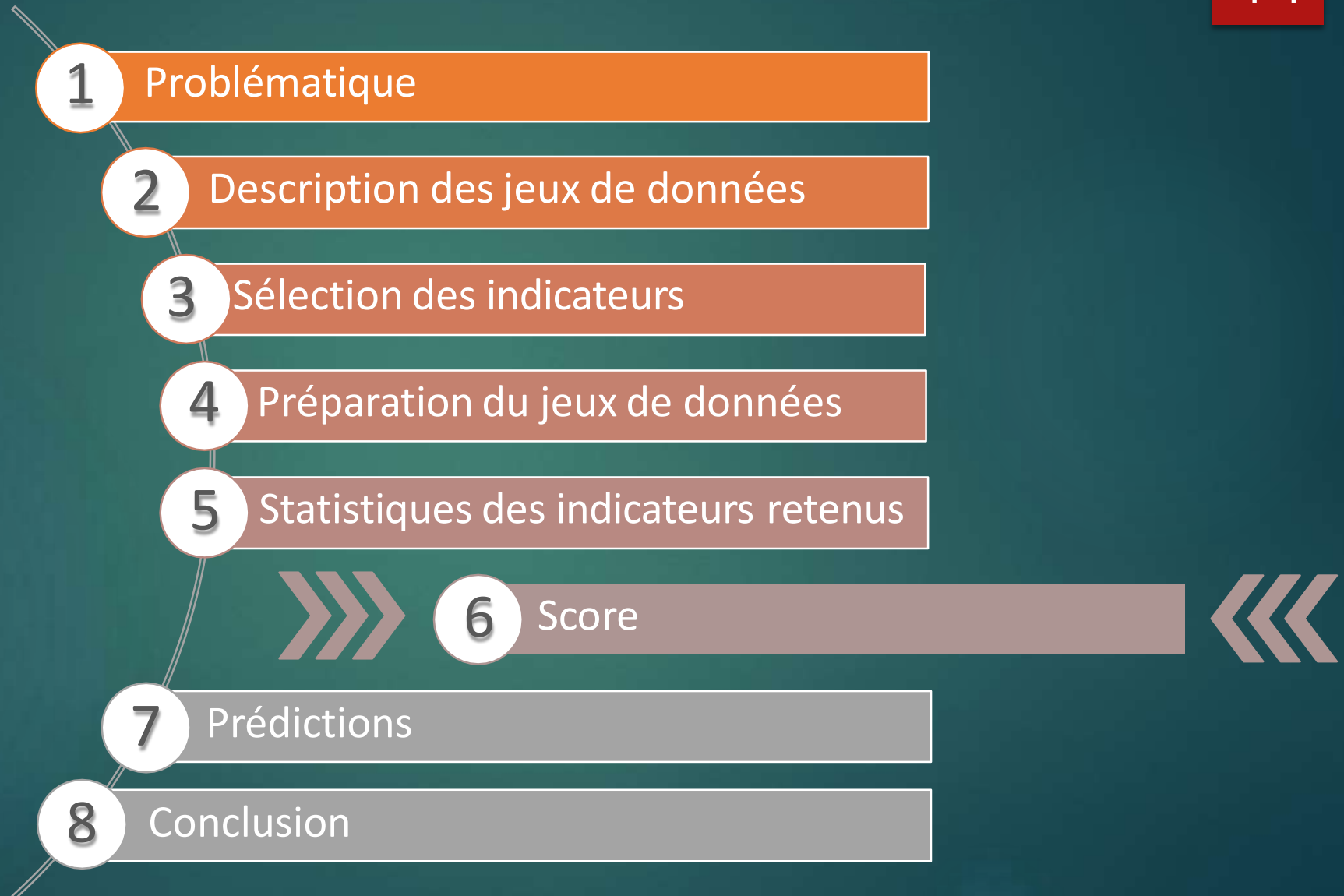


Statistiques des indicateurs retenus

	Description	IT.NET.USER.P2	NY.GNP.PCAP.PP.CD	SE.SEC.ENRR	SP.POP.1524.TO.UN	SE.TER.ENRR
0	mean	40.086642	1.931827e+04	88.155686	45.484345	7.108262e+06
1	median	37.385000	1.263000e+04	92.825211	44.364792	1.130263e+06
2	var	678.174969	3.538400e+08	413.197662	668.473869	7.328145e+14
3	std	26.041793	1.881064e+04	20.327264	25.854862	2.707055e+07
4	skew	0.314415	1.942296e+00	-1.004492	0.213457	7.526472e+00
5	kurtosis	-1.045503	4.536687e+00	1.218952	-0.949891	5.967148e+01

Unités différentes, grandeurs différentes, Comparer ?

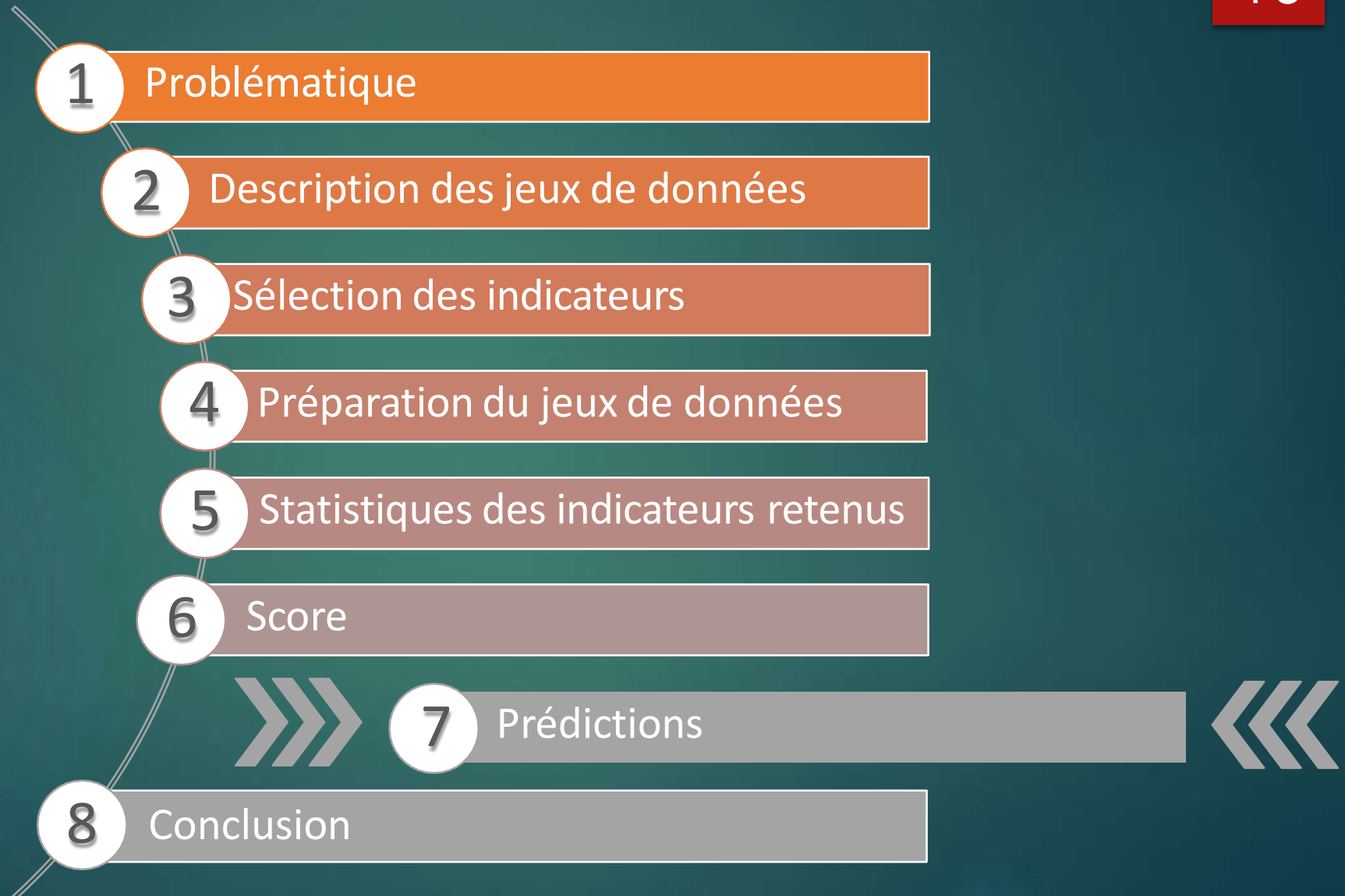
Mise à l'échelle



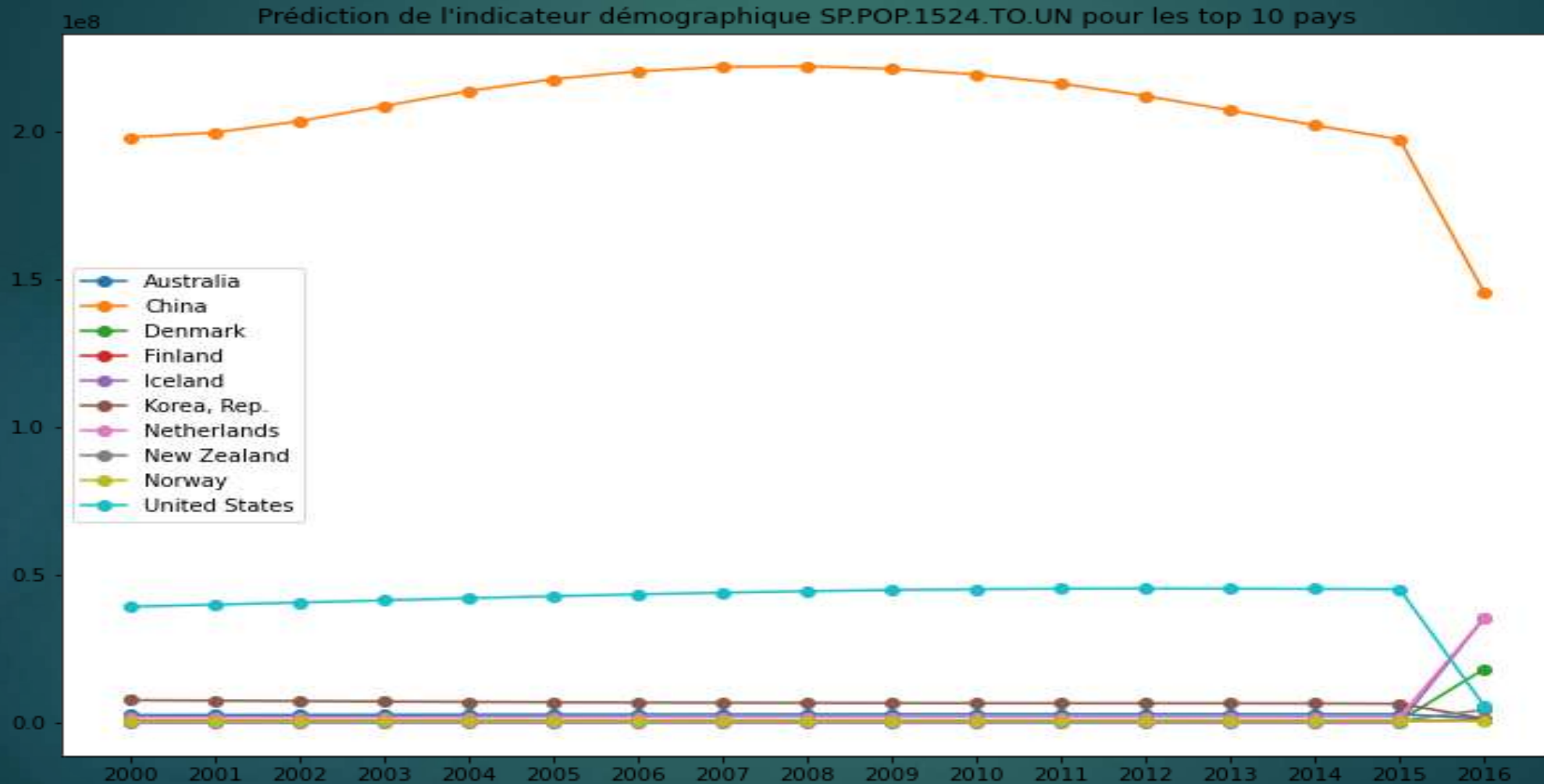
Le score

- 131 pays candidats
- Année choisie :2010
- ... au top 10 des pays attractifs

	Country Name	score	Place Provisoire
123	United States	6.355694	1
86	Norway	6.183803	2
7	Australia	5.797147	3
82	Netherlands	5.576384	4
34	Denmark	5.572011	5
40	Finland	5.519097	6
26	China	5.479156	7
83	New Zealand	4.855338	8
62	Korea, Rep.	4.806124	9
51	Iceland	4.802662	10

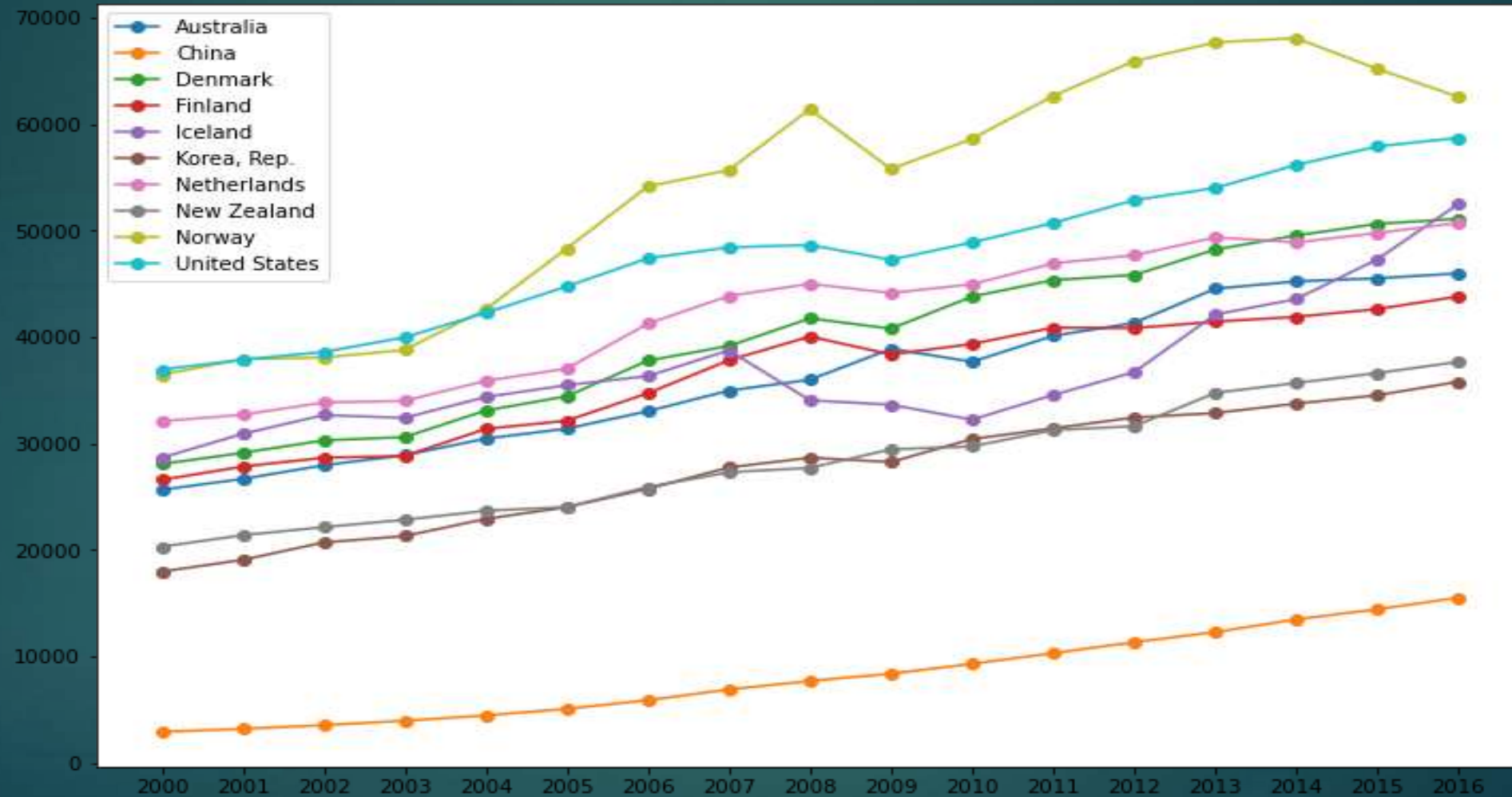


Population



Économie

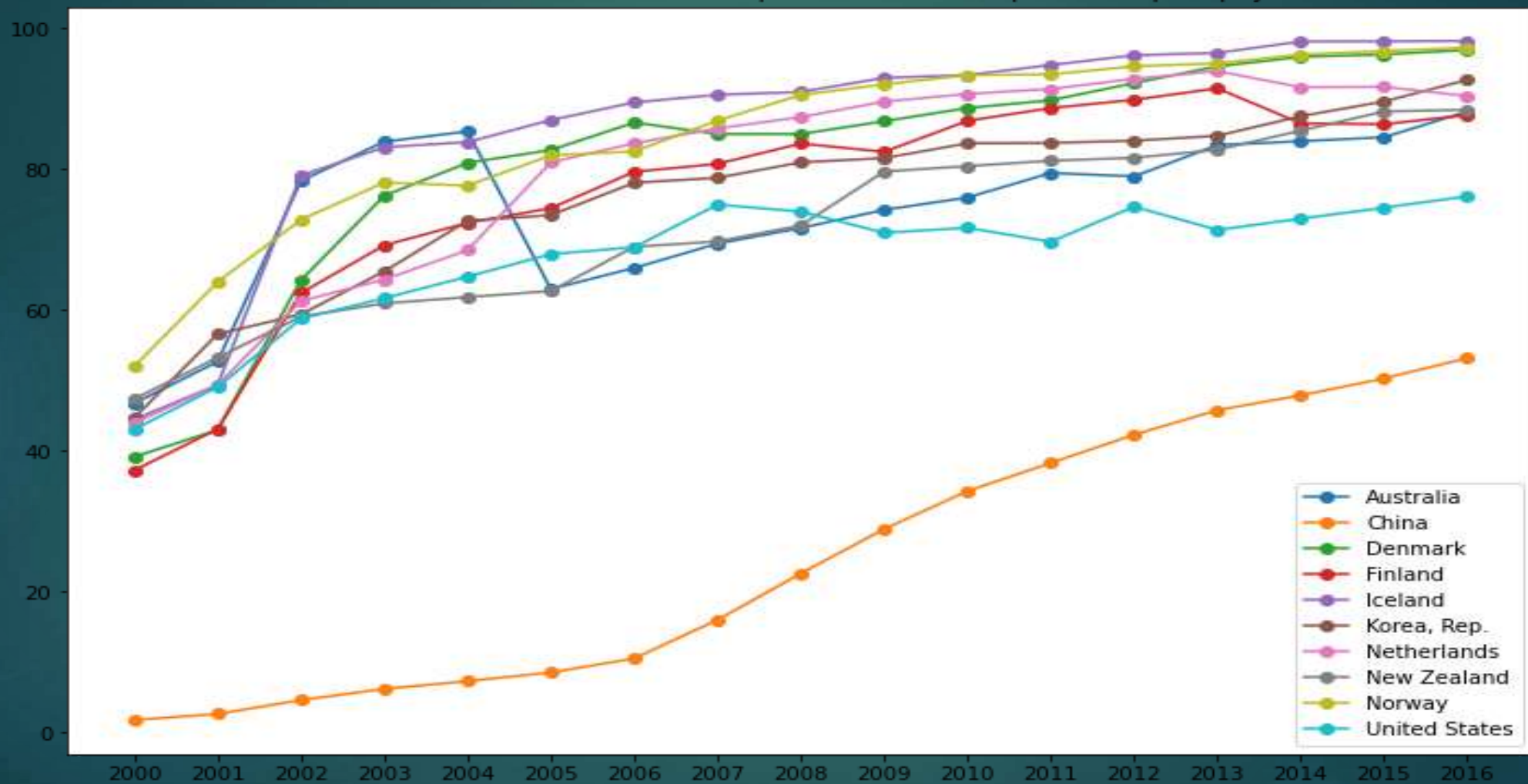
Prédiction de l'indicateur économique NY.GNP.PCAP.PP.CD pour les top 10 pays





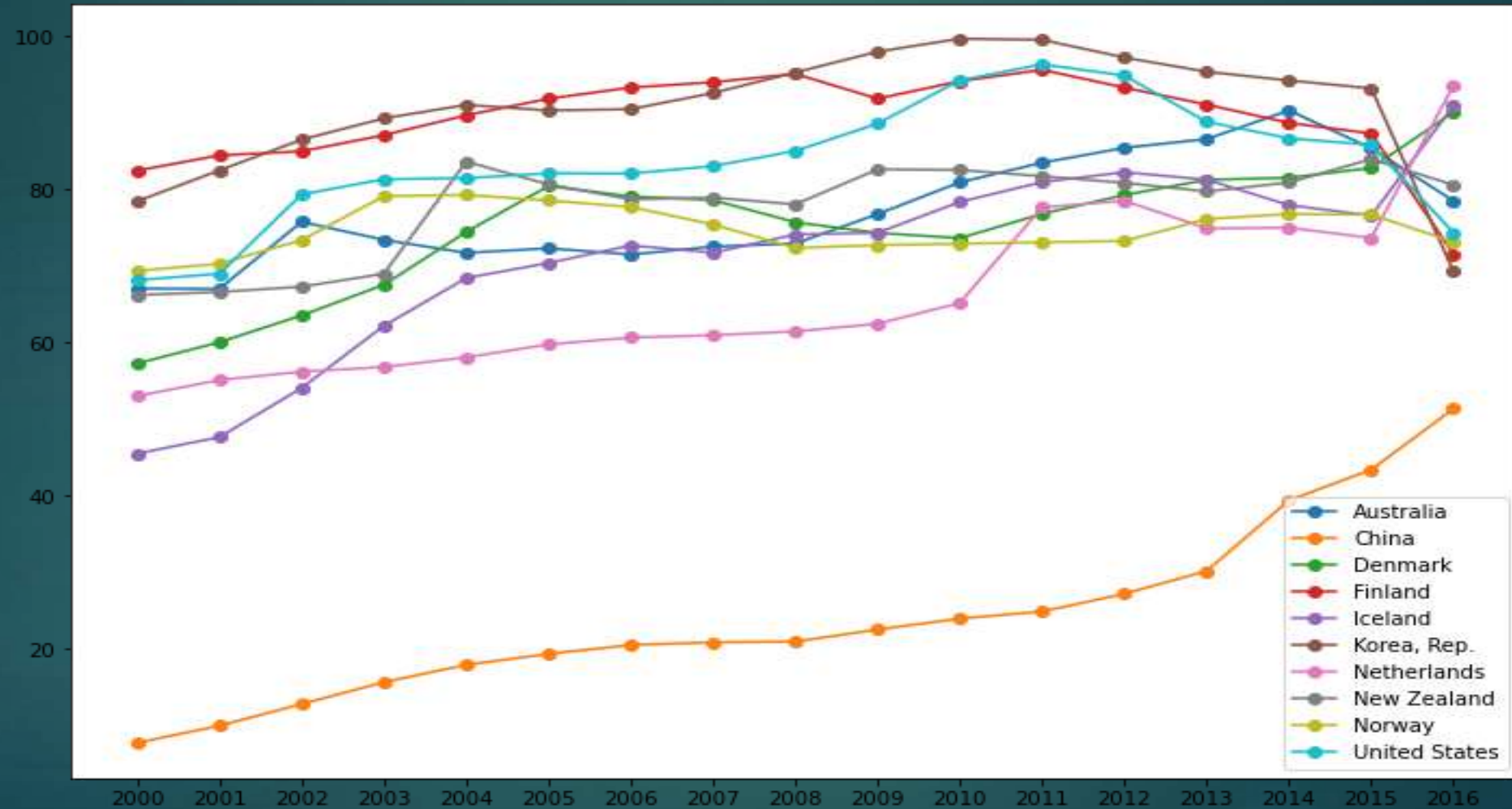
Internet

Prédiction de l'indicateur numérique IT.NET.USER.P2 pour les top 10 pays



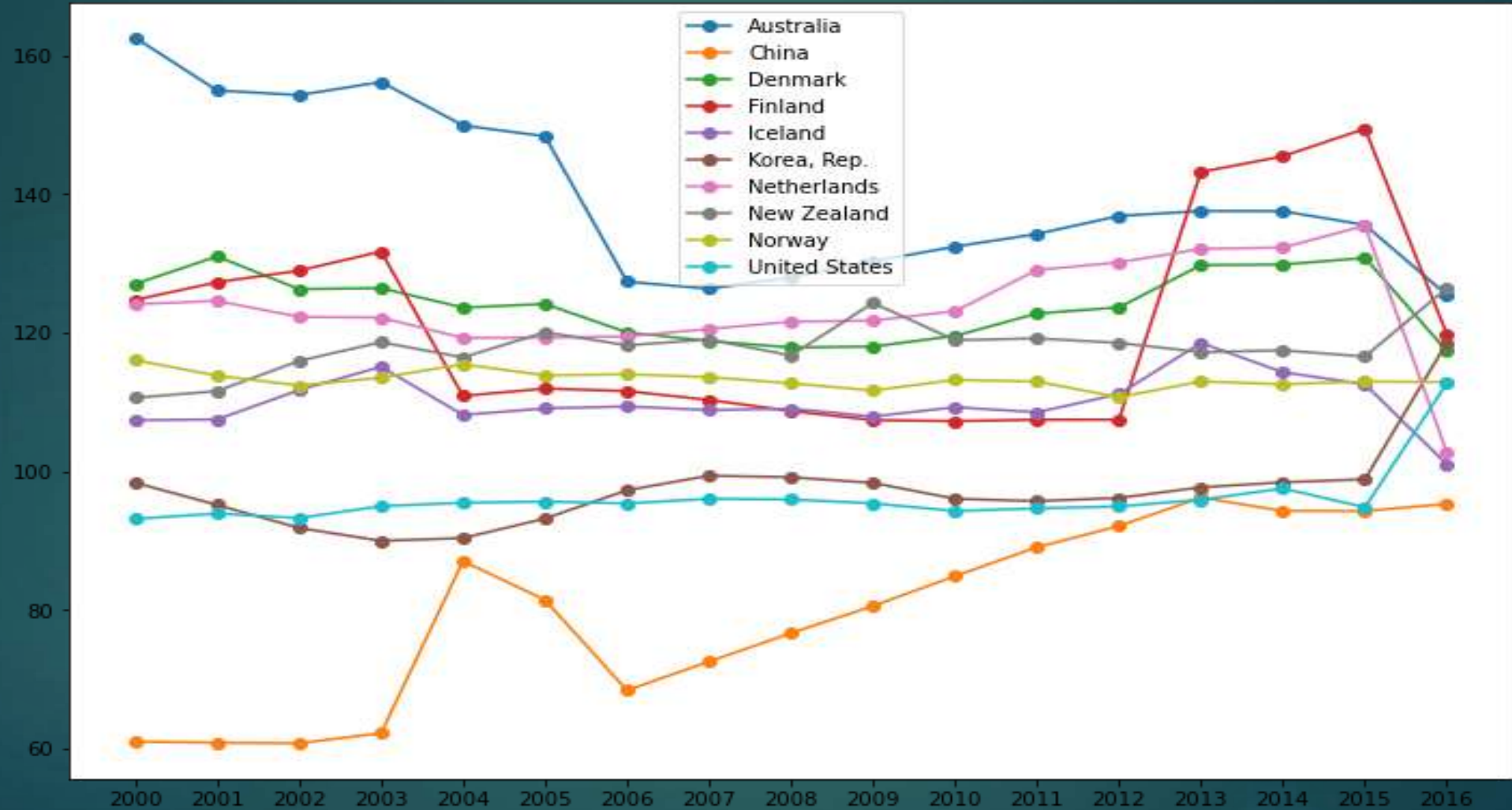
Éducation supérieure

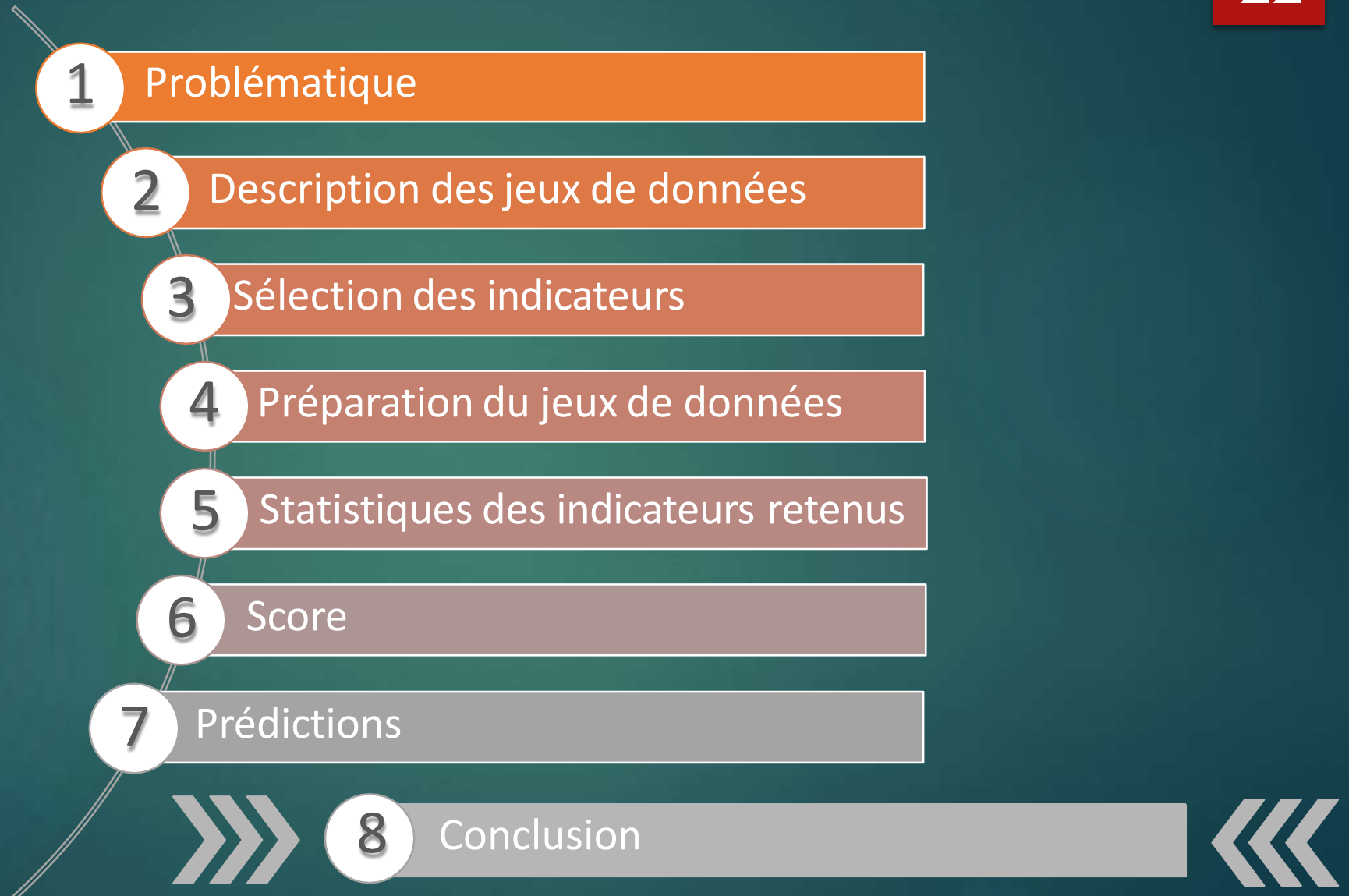
Prédiction de l'indicateur éducatif enseignement supérieur SE.TER.ENRR pour les top 10 pays



Éducation lycée

Prédiction de l'indicateur éducatif lycée SE.SEC.ENRR pour les top 10 pays





Conclusion

- Les pays à fort potentiel sont :
 - ❑ Les Etats Unis, avec un indicateur d'éducation supérieure très élevé
 - ❑ Les pays scandinaves, connus par un système éducatif d'excellence.
 - ❑ Australie, possède un fort taux de scolarisation, et la langue officielle c'est l'anglais.
 - ❑ La Chine, une puissance économique et c'est le pays le plus peuplé au monde
 - ❑ La Corée de Sud, un pays émergent avec d'énormes qualités