

A faint, stylized world map in shades of orange and green serves as the background for the slide.

# Openclassrooms: Projet 2

## Analysez les données de systèmes éducatifs

**Objectif : Comprendre le potentiel client par pays grâce à la base EdStats de la Banque mondiale.**

**Denis MUTOMBO TSHITUKA Mars 2025 Parcours « Data scientist »**



# academy

## Sommaire de la présentation

### 1. Problématique

Les questions clés pour l'internationalisation.

### 2. Jeu de données

Exploration et nettoyage des données.

### 3. Analyse

Sélection et tests statistiques des indicateurs.

### 4. Prédictions

Score et évolution prévisionnelle par pays.



# Problématique du projet



## Questions clés

Identifier les pays à fort potentiel clients et prioriser les marchés.



## Données disponibles

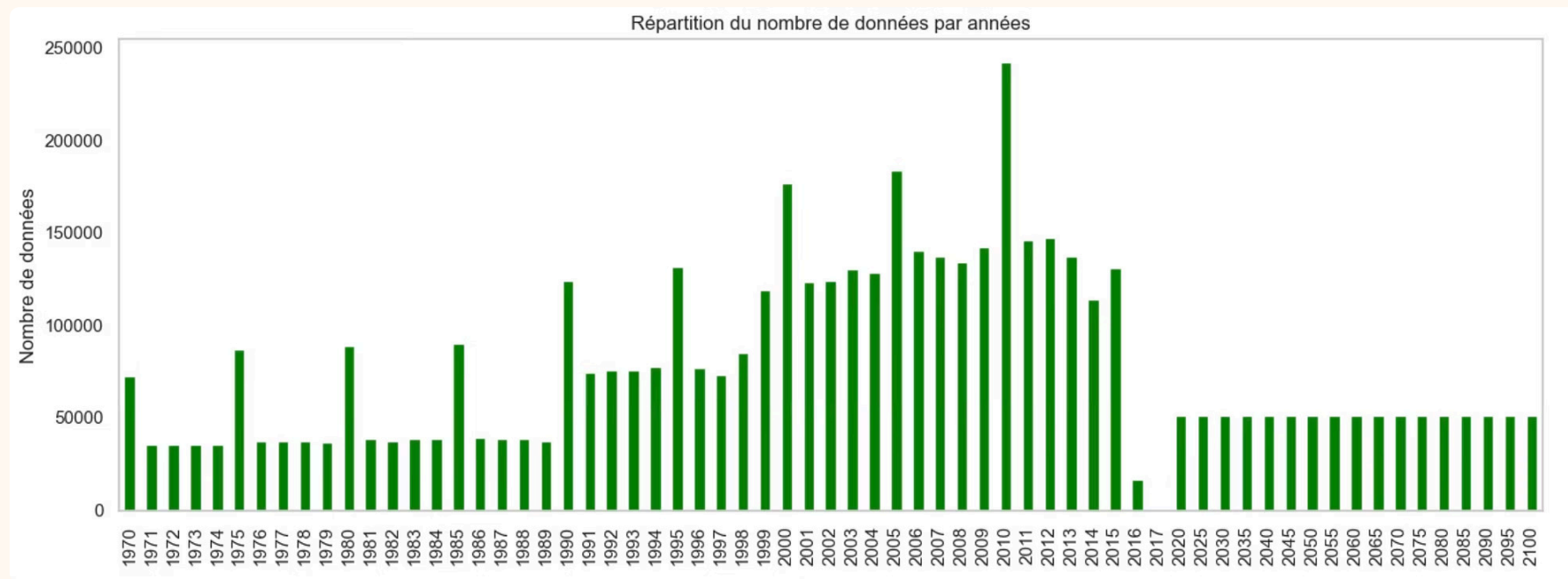
Le portail EdStats de la Banque mondiale offre des données ouvertes sur l'éducation.



## Objectif

Valider la cohérence et tirer des conclusions stratégiques.

# Exploration du jeu de données



## Données brutes

886930 lignes, 70 colonnes de données éducatives de 1970 à 2100.

3665 indicateurs et 242 Pays

## But

Extraire des informations pertinentes pour l'analyse du potentiel marché.

# Réduction du jeu de données

## Données retenues

- Population de 15 à 24 ans, lycéens et universitaires, utilisateurs d'internet
- 2010 : année la plus riche en données, 2015 : la plus proche du présent exploitable.

## Objectif

Identifier des indicateurs pertinents pour orientation stratégique.

# Choix des indicateurs clés

1

SE.SEC.ENRR : Taux de complétion scolaire niveau secondaire & lycée.

2

SE.TER.ENRR : Taux de complétion niveau universitaires.

3

SP.POP.1524.TO.UN : Population 15-24 ans totale.

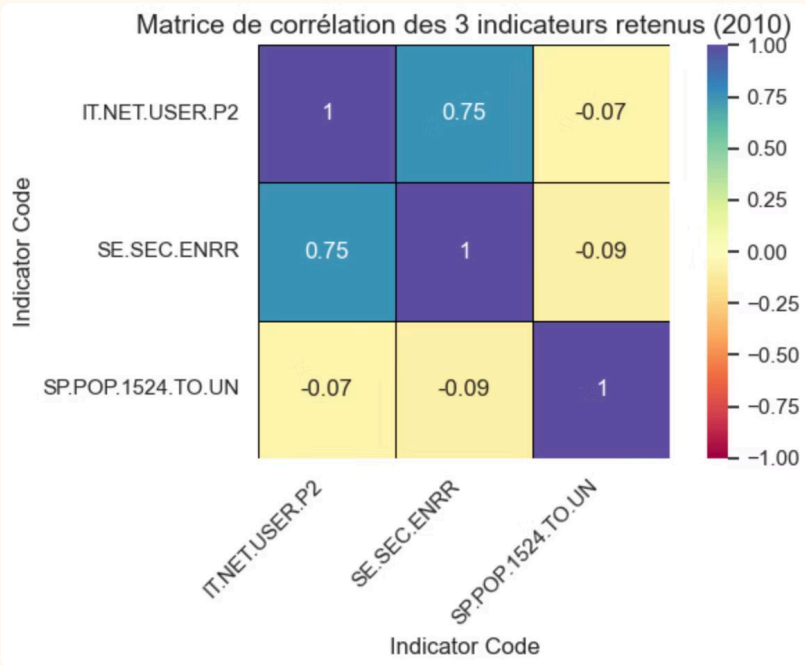
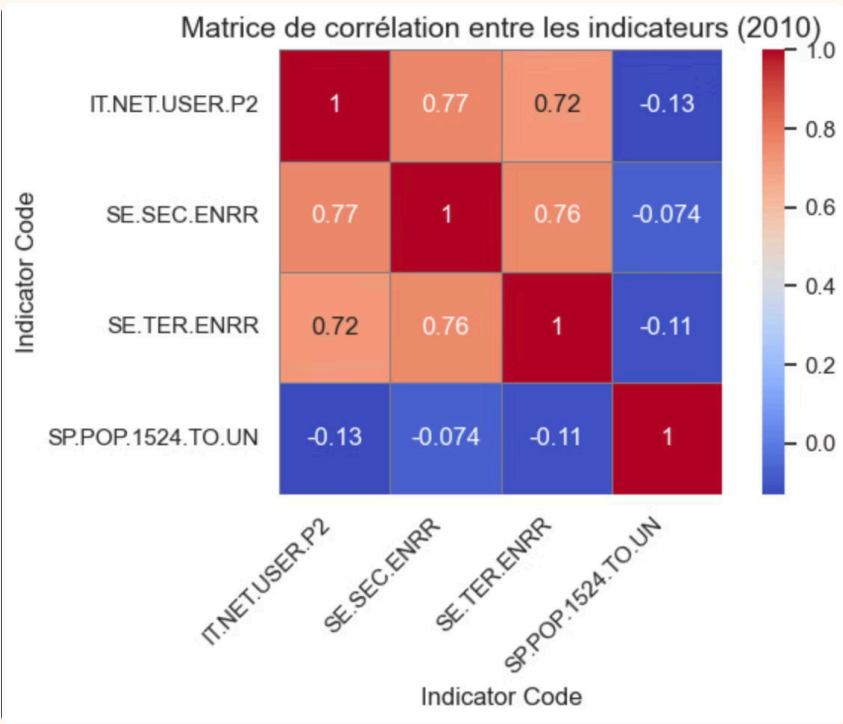
4

IT.NET.USER.P2 : Utilisateurs d'internet (% population).

5

NY.GDP.PCAP.CD : PIB par habitant (USD).

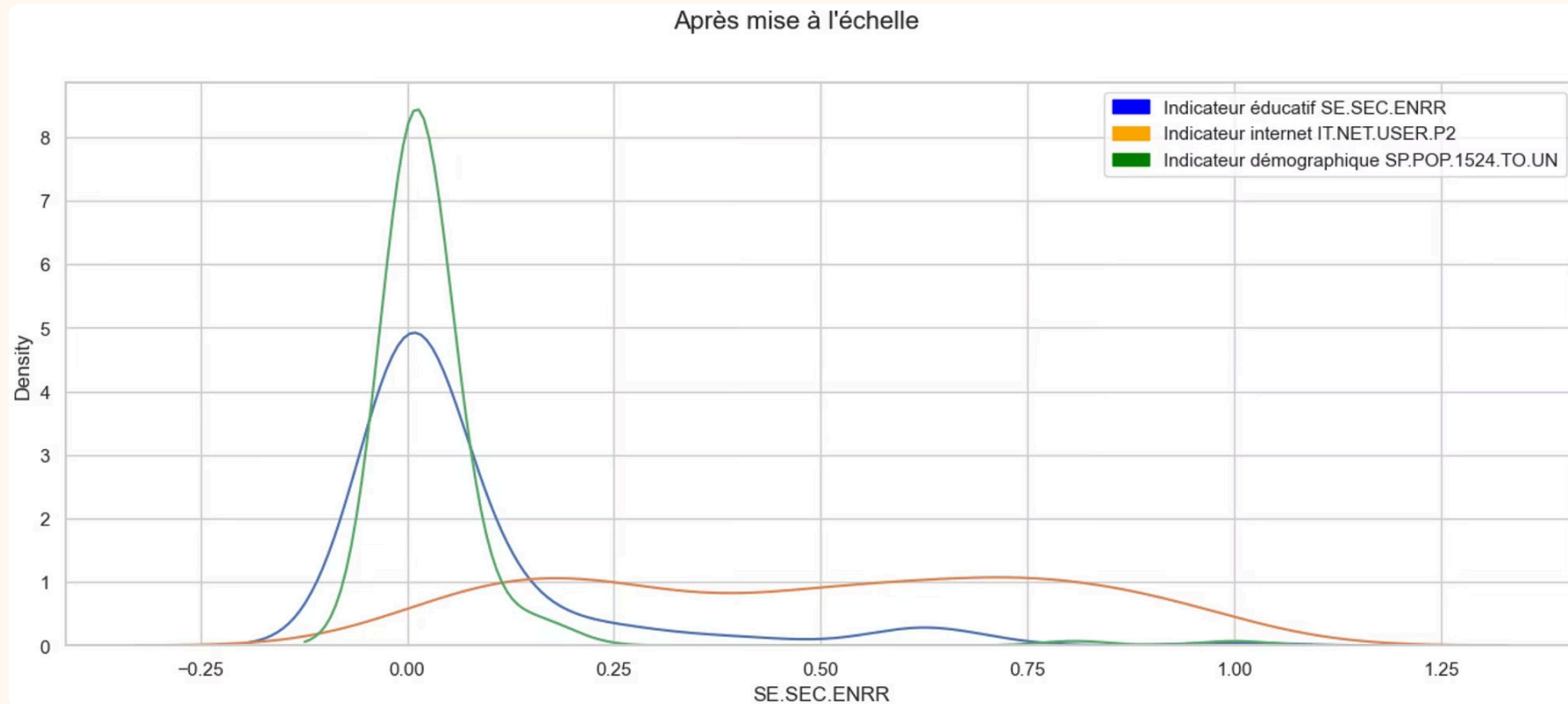
# Analyse statistique approfondie



Matrice de corrélation

La matrice révèle des dépendances entre plusieurs indicateurs.

## Graphique des trois indicateurs clés normalisés à la même échelle





# Scoring des pays candidats

## Pays retenus

1

Sur 140 candidats, 10 pays sélectionnés par score global.

Australie, Japon, États-unis, France, Pologne,  
Luxembourg, Royaume-Uni, Islande, Slovénie,  
Arabie saoudite.

2

## Zones géographiques clés

Quatre zones se démarquent fortement.

Est Asie & Pacifique, Amérique du nord, Europe &  
Asie centrale, Moyen-Orient.

Carte des scores moyens (éducation + numérique)



# Prédictions

1

## Augmentation

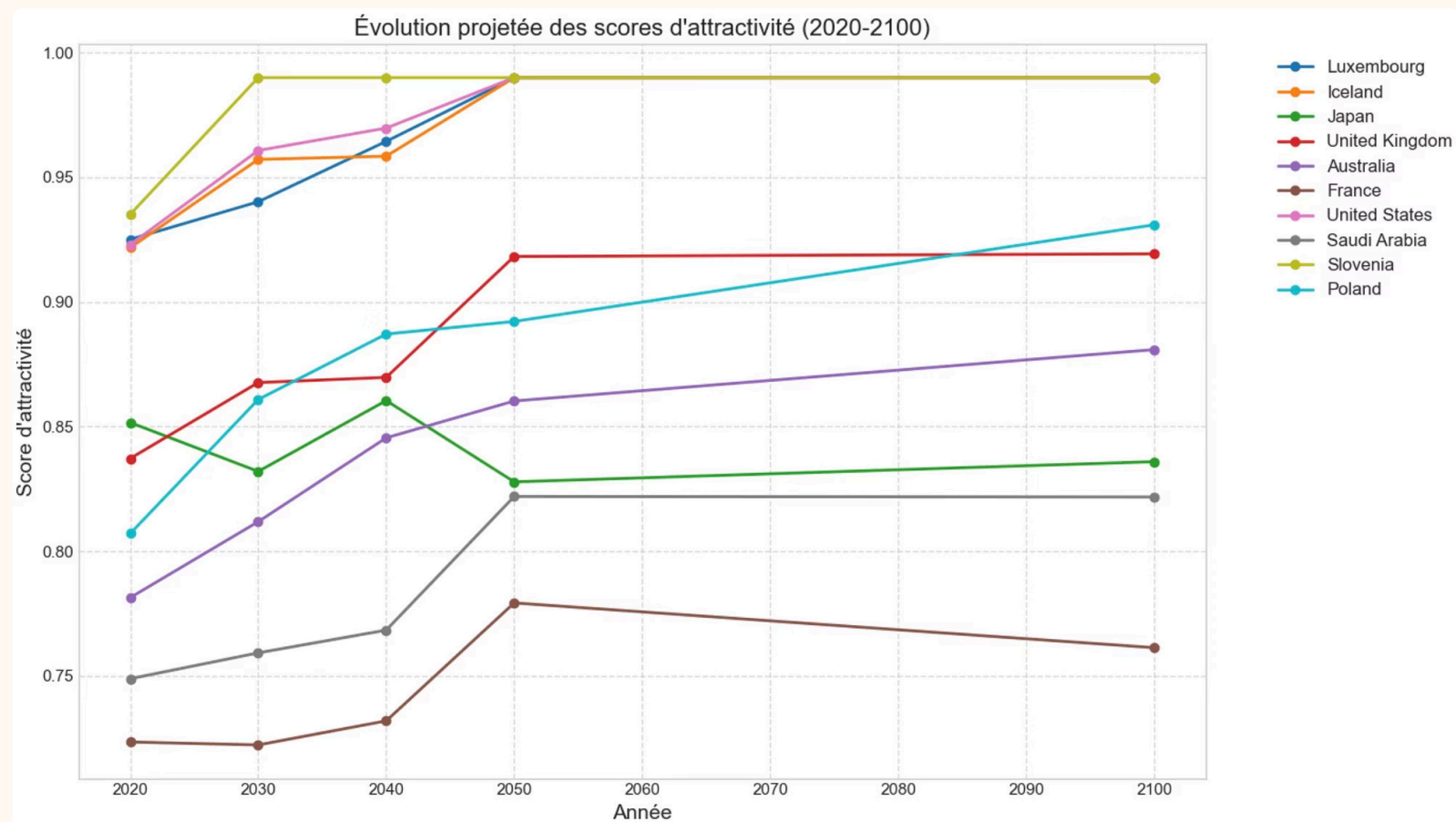
Certains pays montrent une croissance importante des 15-24 ans.

2

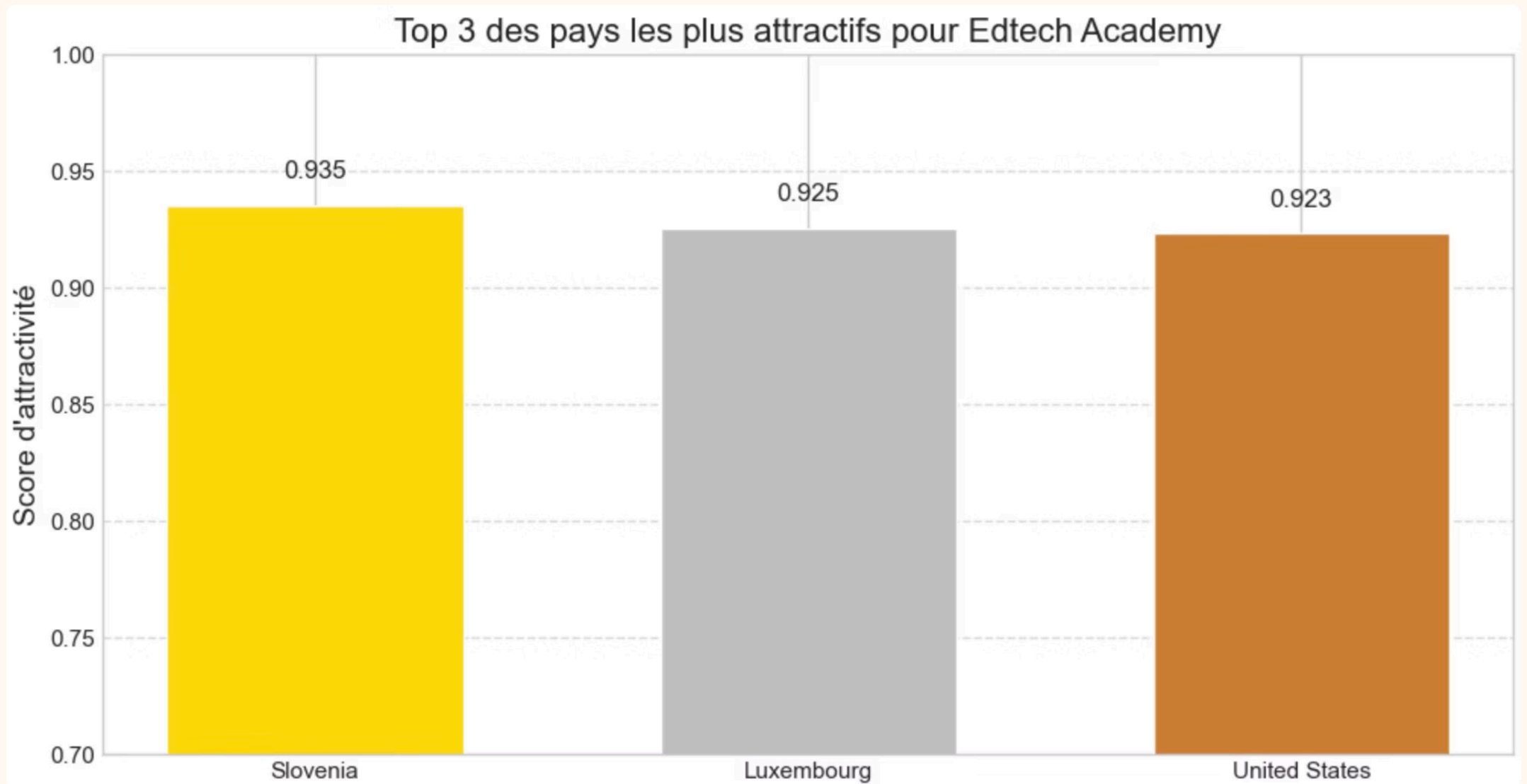
## Diminution

D'autres pays présentent un recul de cette population clé.

# Score d'attractivité de 2020-2100



# Conclusions



## Pays prioritaires

- Slovénie
- Luxembourg
- États-Unis