

# Projet 2 : Analyse de données de systèmes éducatifs

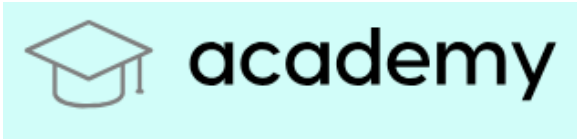
Fanjamalala Rajaonalison  
29/04/2021



# Plan

---

1. Présentation de l'étude
2. Présentation des données
3. Exploration des données
4. Tri des données
5. Analyses comparatifs



- Start-up de la EdTech
- Formation en ligne pour un public de niveau lycée et université
- Projet d'expansion à l'international



- Start-up de la EdTech
- Formation en ligne pour un public de niveau lycée et université
- Projet d'expansion à l'international



- Pays avec un fort potentiel de clients ?
- Evolution de ce potentiel de clients ?
- Priorité des pays ?



- Start-up de la EdTech
- Formation en ligne pour un public de niveau lycée et université
- Projet d'expansion à l'international



- Pays avec un fort potentiel de clients ?
- Evolution de ce potentiel de clients ?
- Priorité des pays ?



Déterminer si les données permettent d'informer le projet d'expansion :

- Identifier les données pertinentes
- Comparer les données par pays / régions
- Prioriser les pays / régions selon leurs potentiels

Données de la Banque Mondiale :

- Plus de 4 000 indicateurs comparables au niveau international
- Informations sur l'éducation, la population, les dépenses, les ressources, etc.
- 5 fichiers .csv

1 numériques  
(EdStatsData)

textes

| Country Name | Country Code | Indicator Name                                    | Indicator Code | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | ... | 2060 | 2065 | 2070 | 2075 | 2080 | 2085 | 2090 | 2095 | 2100 |
|--------------|--------------|---|----------------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|------|------|
| Arab World   | ARB          | Adjusted net enrolment rate, lower secondary, ... | UIS.NERA.2     | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | ... | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |

| Country Code | Short Name | Table Name | Long Name | 2-alpha code | Currency Unit | Special Notes                                     | Region                    | Income Group         | WB-2 code | ... | IMF data dissemination standard | Latest population census | Latest household survey | Source of most recent income and expenditure data |
|--------------|------------|------------|-----------|--------------|---------------|---|---------------------------|----------------------|-----------|-----|---------------------------------|--------------------------|-------------------------|---|
| ABW          | Aruba      | Aruba      | Aruba     | AW           | Aruban florin | SNA data for 2000-2011 are updated from offici... | Latin America & Caribbean | High income: nonOECD | AW        | ... | NaN                             | 2010                     | NaN                     | NaN   |

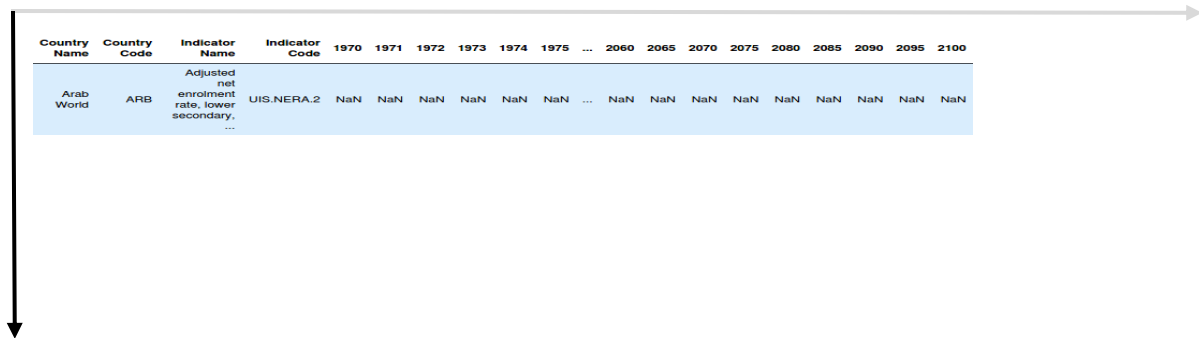
  

| Series Code         | Topic      | Indicator Name                                    | Short definition                                  | Long definition                                   | Unit of measure | Periodicity | Base Period | Other notes | Aggregation method | ... | Notes from original source | General comments |
|---------------------|------------|---|---|---|-----------------|-------------|-------------|-------------|--------------------|-----|----------------------------|------------------|
| BAR.NOED.1519.FE.ZS | Attainment | Barro-Lee: Percentage of female population age... | Percentage of female population age 15-19 with... | Percentage of female population age 15-19 with... | NaN             | NaN         | NaN         | NaN         | NaN                | ... | NaN                        | NaN              |

| Country Name | Country Code | Indicator Name                                    | Indicator Code | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | ... | 2060 | 2065 | 2070 | 2075 | 2080 | 2085 | 2090 | 2095 | 2100 |
|--------------|--------------|---|----------------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|------|------|
| Arab World   | ARB          | Adjusted net enrolment rate, lower secondary, ... | UIS.NERA.2     | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | ... | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |

Colonnes :

- N = 70 dont 1 vide
- 4 colonnes descriptives
- 65 colonnes numériques (Années)
- Beaucoup de valeurs manquantes :  
72,66% → 99,98%



| Country Name | Country Code | Indicator Name                                    | Indicator Code | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | ... | 2060 | 2065 | 2070 | 2075 | 2080 | 2085 | 2090 | 2095 | 2100 |
|--------------|--------------|---|----------------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|------|------|
| Arab World   | ARB          | Adjusted net enrolment rate, lower secondary, ... | UIS.NERA.2     | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | ... | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |

Colonnes :

- N = 70 dont 1 vide
- 4 colonnes descriptives
- 65 colonnes numériques (Années)
- Beaucoup de valeurs manquantes : 72,66% → 99,98%

Lignes :

- N = 886930 lignes dont 529525 lignes vides
- Beaucoup de valeurs manquantes : 27,69% → 98,46%
- Pas de valeurs dupliquées
- 3665 Indicateurs
- 242 Pays



Suppression des lignes complètement vides

- 3665 indicateurs
- 242 pays

65 années

| Country Name | Country Code | Indicator Name                                    | Indicator Code | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | ... | 2060 | 2065 | 2070 | 2075 | 2080 | 2085 | 2090 | 2095 | 2100 |
|--------------|--------------|---|----------------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|------|------|
| Arab World   | ARB          | Adjusted net enrolment rate, lower secondary, ... | UIS.NERA.2     | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | ... | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |

Suppression des lignes complètement vides



Sélection des pays selon leur nombre d'indicateurs



Sélection des indicateurs selon leur nombre d'indicateurs

$N_{\text{indicateurs}} > \text{distribution basse}$

$N_{\text{pays}} > \text{distribution basse}$

Suppression des lignes complètement vides



Sélection des pays selon leur nombre d'indicateurs



Sélection des indicateurs selon leur nombre d'indicateurs



Tri des indicateurs en fonction de leur topic



Choix de 10 indicateurs parmi les topics sélectionnés



Sélection des pays ayant des données pour tous les 10  
indicateurs uniquement



Sélection des grands pays (pays ayant un nombre élevé de  
population)

Topics relatifs à 4 axes : DEMOGRAPHIE, EDUCATION,  
RESSOURCE MATERIELLE, RESSOURCE ECONOMIQUE

$N_{\text{population}} > \text{médiane}$

Suppression des lignes complètement vides



Sélection des pays selon leur nombre d'indicateurs



Sélection des indicateurs selon leur nombre d'indicateurs



Tri des indicateurs en fonction de leur topic



Choix de 10 indicateurs parmi les topics sélectionnés



Sélection des pays ayant des données pour tous les 10  
indicateurs uniquement



Sélection des grands pays (pays ayant un nombre élevé de  
population)



Sélection d'années

Années récentes

Suppression des lignes complètement vides



Sélection des pays selon leur nombre d'indicateurs



Sélection des indicateurs selon leur nombre d'indicateurs



Tri des indicateurs en fonction de leur topic



Choix de 10 indicateurs parmi les topics sélectionnés



Sélection des pays ayant des données pour tous les 10 indicateurs uniquement



Sélection des grands pays (pays ayant un nombre élevé de population)



Sélection d'années

- 3665 indicateurs
- 242 pays

65 années

| Country Name | Country Code | Indicator Name                                    | Indicator Code | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | ... | 2060 | 2065 | 2070 | 2075 | 2080 | 2085 | 2090 | 2095 | 2100 |
|--------------|--------------|---|----------------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|------|------|
| Arab World   | ARB          | Adjusted net enrolment rate, lower secondary, ... | UIS.NERA.2     | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | ... | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |

- 10 indicateurs
- 61 pays

22 années

| Country Name | Country Code | Indicator Name                                    | Indicator Code | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | ... | 2060 | 2065 | 2070 | 2075 | 2080 | 2085 | 2090 | 2095 | 2100 |
|--------------|--------------|---|----------------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|------|------|
| Arab World   | ARB          | Adjusted net enrolment rate, lower secondary, ... | UIS.NERA.2     | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | ... | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  | NaN  |





## 6 pays à fort potentiel

| Pays          | personnes en âge de travail (10 <sup>6</sup> ) | 15+ ans avec niveau lycée ou université (%) | utilisateurs d'Internet (%) | RNB par habitant (international \$) |
|---------------|--|---|-----------------------------|-------------------------------------|
| United States | 152  | 48.36                                       | 57.22                       | 44066.82                            |
| Japan         | 67   | 41.22                                       | 56.20                       | 32603.18                            |
| Netherlands   | 8  | 42.58                                       | 65.88                       | 38557.73                            |
| Canada        | 18   | 45.67                                       | 61.91                       | 34528.18                            |
| Australia     | 11   | 44.52                                       | 58.56                       | 32997.27                            |
| Germany       | 41   | 39.33                                       | 56.36                       | 35054.09                            |

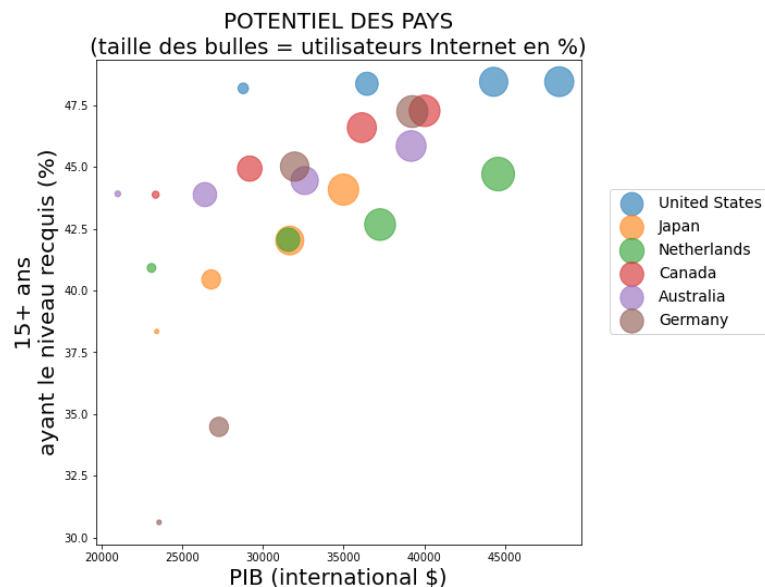
USA : potentiellement le plus intéressant

- Forte économie
- Grande taille population cible
- Ressource adaptée



## 6 pays à fort potentiel

| Pays          | personnes en âge de travail ( $10^6$ ) | 15+ ans avec niveau lycée ou université (%) | utilisateurs d'Internet (%) | RNB par habitant (international \$) |
|---------------|--|---|-----------------------------|-------------------------------------|
| United States | 152                                    | 48.36                                       | 57.22                       | 44066.82                            |
| Japan         | 67                                     | 41.22                                       | 56.20                       | 32603.18                            |
| Netherlands   | 8                                      | 42.58                                       | 65.88                       | 38557.73                            |
| Canada        | 18                                     | 45.67                                       | 61.91                       | 34528.18                            |
| Australia     | 11                                     | 44.52                                       | 58.56                       | 32997.27                            |
| Germany       | 41                                     | 39.33                                       | 56.36                       | 35054.09                            |

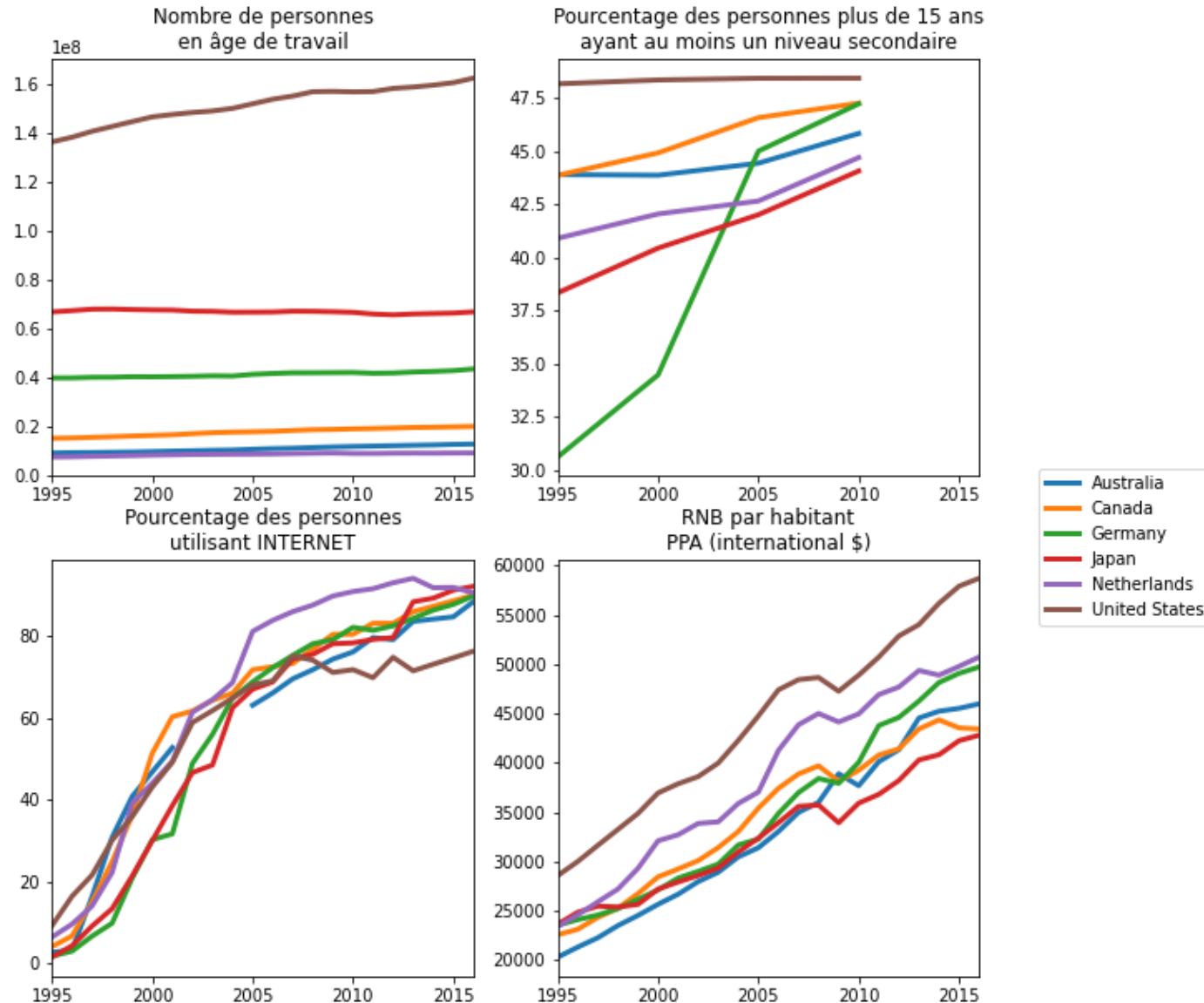


### USA : potentiellement le plus intéressant

- Forte économie
- Grande taille population cible
- Ressource adaptée

Education semble être influencée par les ressources (économie et matériel)

## Evolution des 6 pays

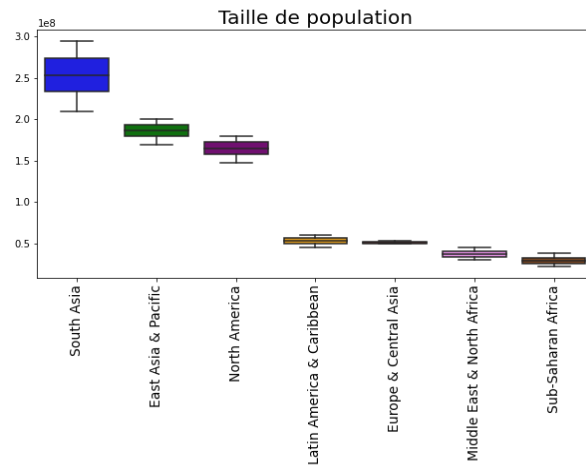


- Courbes à tendance croissante
- Pas énormément de différence pour l'utilisation d'Internet
- USA : presque toujours le plus intéressant
- Germany : Forte croissance de l'éducation et de l'économie → bonne évolution du potentiel

# Analyses comparatifs

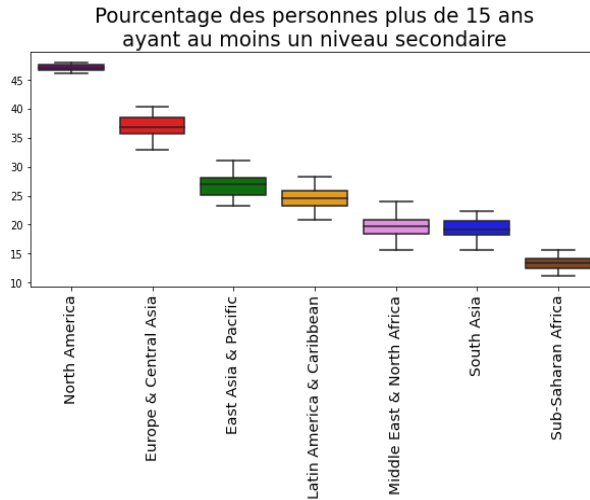
## Par zones géographiques

### DEMOGRAPHIE



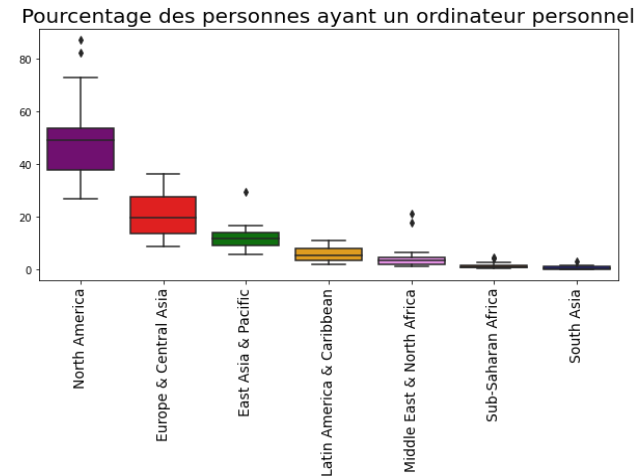
South Asia  
East Asia & Pacific  
North America

### EDUCATION



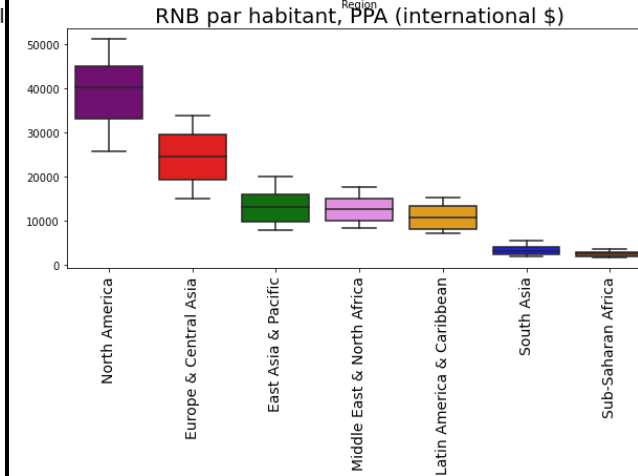
North America  
Europe & Central Asia  
East Asia & Pacific

### RESSOURCE MATERIELLE



North America  
Europe & Central Asia  
East Asia & Pacific

### RESSOURCE ECONOMIQUE



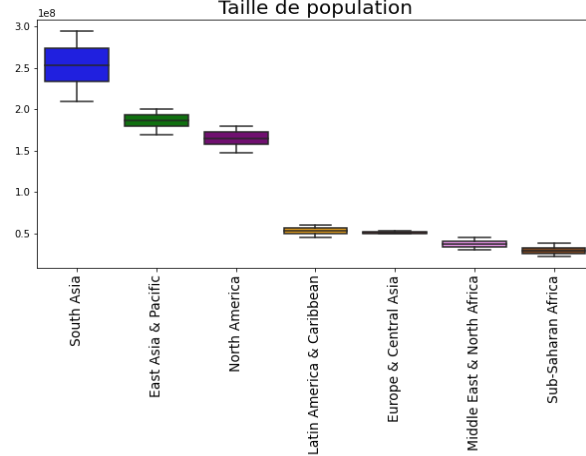
North America  
Europe & Central Asia  
East Asia & Pacific

# Analyses comparatifs

## Par zones géographiques

### DEMOGRAPHIE

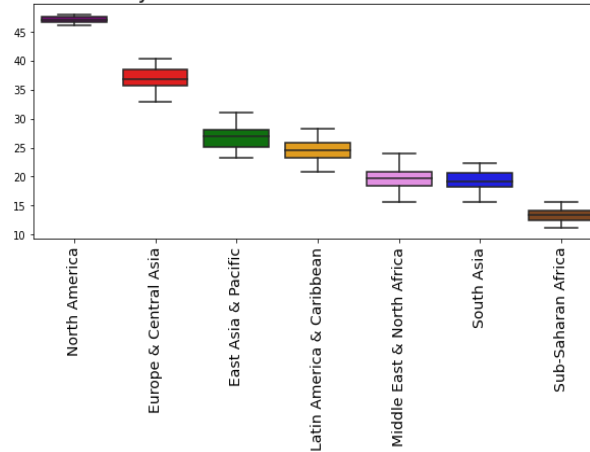
Taille de population



South Asia  
East Asia & Pacific \*\*\*\*  
North America \*\*\*\*

### EDUCATION

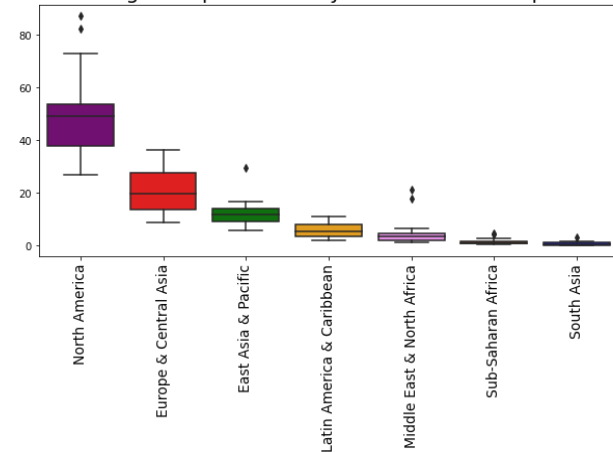
Pourcentage des personnes plus de 15 ans  
ayant au moins un niveau secondaire



North America \*\*\*\*  
Europe & Central Asia +++  
East Asia & Pacific \*\*\*\*

### RESSOURCE MATERIELLE

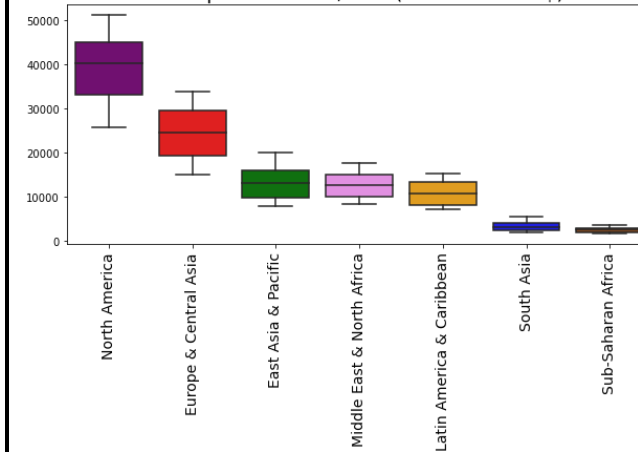
Pourcentage des personnes ayant un ordinateur personnel



North America \*\*\*\*  
Europe & Central Asia +++  
East Asia & Pacific \*\*\*\*

### RESSOURCE ECONOMIQUE

RNB par habitant, PPA (international \$)



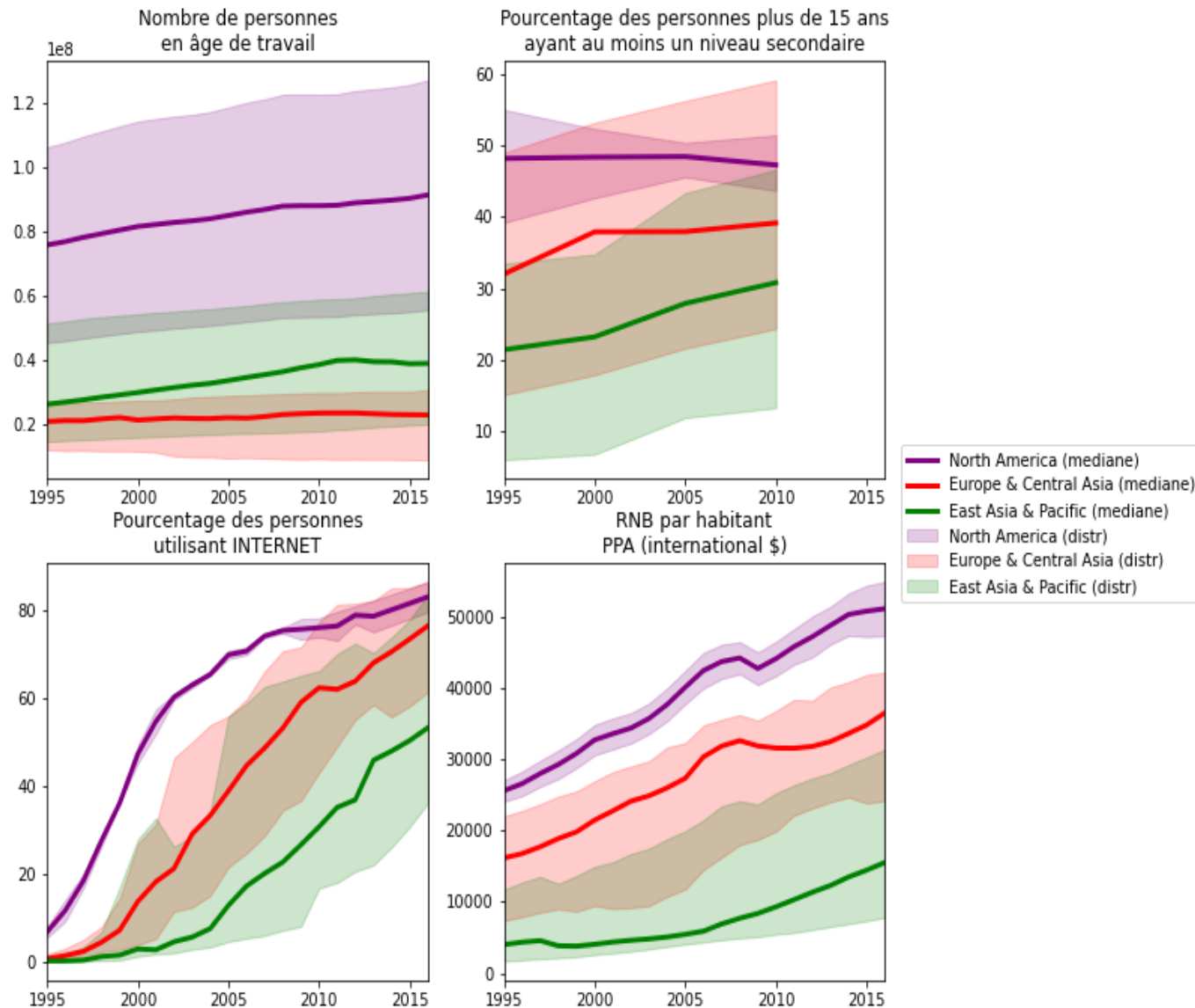
North America \*\*\*\*  
Europe & Central Asia +++  
East Asia & Pacific \*\*\*\*

## 3 zones géographiques à fort potentiel

| Zones géographiques   | personnes en âge de travail (10 <sup>6</sup> ) | 15+ ans avec niveau lycée ou université (%) | utilisateurs d'Internet (%) | RNB par habitant (international \$) |
|-----------------------|--|---|-----------------------------|-------------------------------------|
| North America         | 84.7   | 47.01                                       | 59.56                       | 39297.50                            |
| Europe & Central Asia | 24   | 36.85                                       | 39.59                       | 24363.25                            |
| East Asia & Pacific   | 102  | 26.90                                       | 27.40                       | 13116.28                            |

North America : potentiellement le plus intéressant

## Evolution des 3 zones géographiques



- Courbes à tendance croissante
- Forte variabilité (globalement plus remarquée en East Asia & Pacific)
- North America potentiellement le plus intéressant
- Evolution intéressante de Europe & Central Asia basée sur l'économie



## Conclusion

- ✓ Information sur le projet d'expansion
- ✓ Potentiel actuel :  
North America  
USA
- ✓ Potentiel en évolution :  
Europe & Central Asia  
Germany

---

# MERCI

---

Questions



Incompréhensions

