



EdTech
France

Analyse des données de systèmes éducatifs

Data Science | Projet 2

Firat Yasar
19/06/2021

Sommaire

Présentation

- Présentation de la problématique
- Découverte du jeu de données
- Le jeu de données (Présélection)

Analyse Exploratoire

- RNB (Revenu national brut)
- Internet (Le taux d'utilisation d'Internet)
- Population (Le nombre d'inscriptions dans l'enseignement secondaire)
- Méthode d'analyse

Les pays où l'entreprise doit opérer en priorité



Présentation

Présentation de la problématique

START-UP !



- **L'objectif :**

Le projet d'expansion à l'international de l'entreprise.

Notre **start-up de la EdTech**, nommée **academy**, propose des contenus de formation en ligne pour un public de niveau lycée et université.

- **La mission:**

Une **première mission d'analyse exploratoire**, pour déterminer si les données sur l'éducation de la banque mondiale permettent d'informer le projet d'expansion.

- 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 ?

- **La base de données :**

Pour cela, nous avons à notre disposition la base de données :

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

Découverte du jeu de données : Les données de la Banque mondiale

The screenshot shows the World Bank Data Catalog interface. At the top, there is a search bar and navigation links for 'Data' and 'Collect'. The main heading is 'Education Statistics'. Below this, there is a section for 'DATA ACCESS AND LICENSING' which states the dataset is 'Public' and licensed under 'CC-BY 4.0'. A Creative Commons license icon is displayed. To the right, there are social sharing buttons for print, email, tweet, Facebook share, LinkedIn share, and a plus sign. Below the sharing buttons is a descriptive paragraph about the 'The World Bank EdStats All Indicator Query' which holds over 4,000 indicators. At the bottom, there are tabs for 'Overview', 'Data & Resources', 'Additional Information', and 'Citations'. Under the 'Overview' tab, there is a table with the following information:

| | |
|----------------------|-------------------|
| Type: | Time Series |
| Acronym: | EdStats |
| Languages Supported: | English |
| Topics: | Education, Gender |

Data

886 930 lignes 70 colonnes

3 665 indicateurs

Data : disponible à partir des années
1970

Nombre de pays : 217
+ régions + continents : 242

Le jeu de données (Présélection)



- **3 indicateurs sélectionnés:**

- Population (Inscriptions dans l'enseignement secondaire),
- RNB (Revenu national brut),
- Internet (Le taux d'utilisation d'Internet)

Entre les années 2008-2020 nous avons pris les données les plus récentes

651 lignes × 17 colonnes

Nous avons analysé les 179 pays.

Note (le taux de lignes vides) : sur 651 lignes

pour 2020 : il y a 3 valeurs

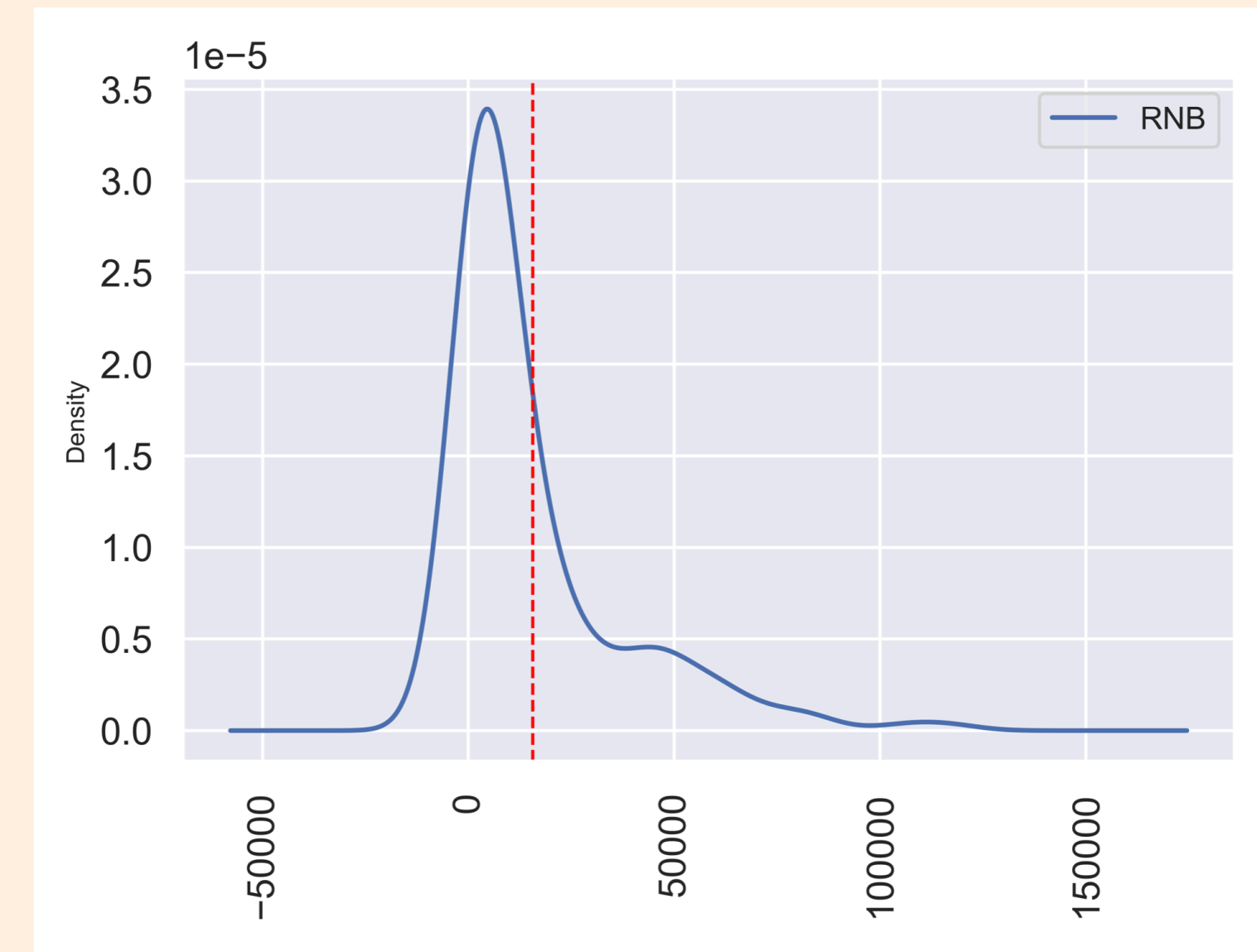
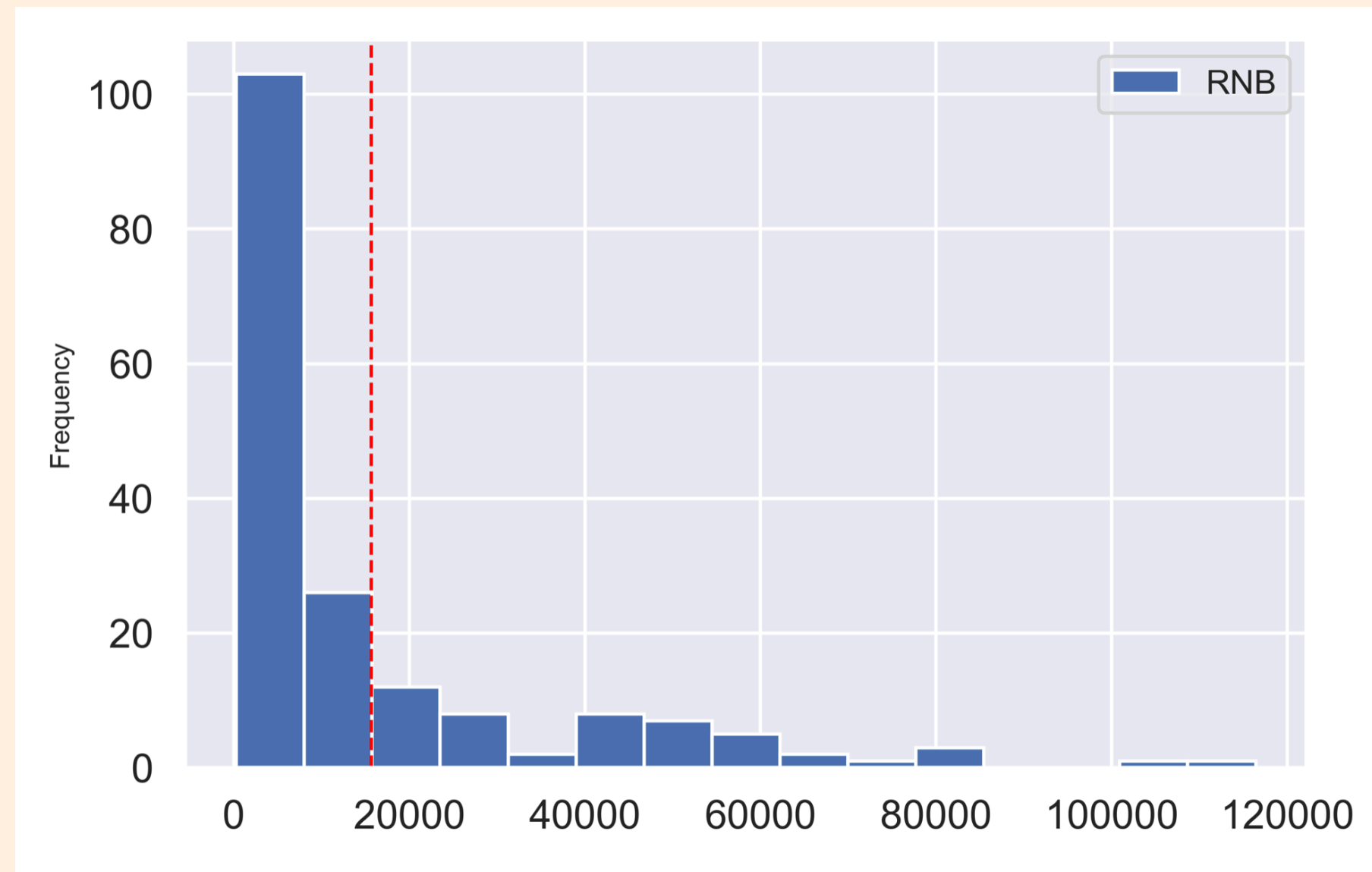
pour 2019 : il y a 67 valeurs

pour 2018 : il y a 388 valeurs

les colonnes 2008-2017 : 17 % - 19 % vides (nuls)

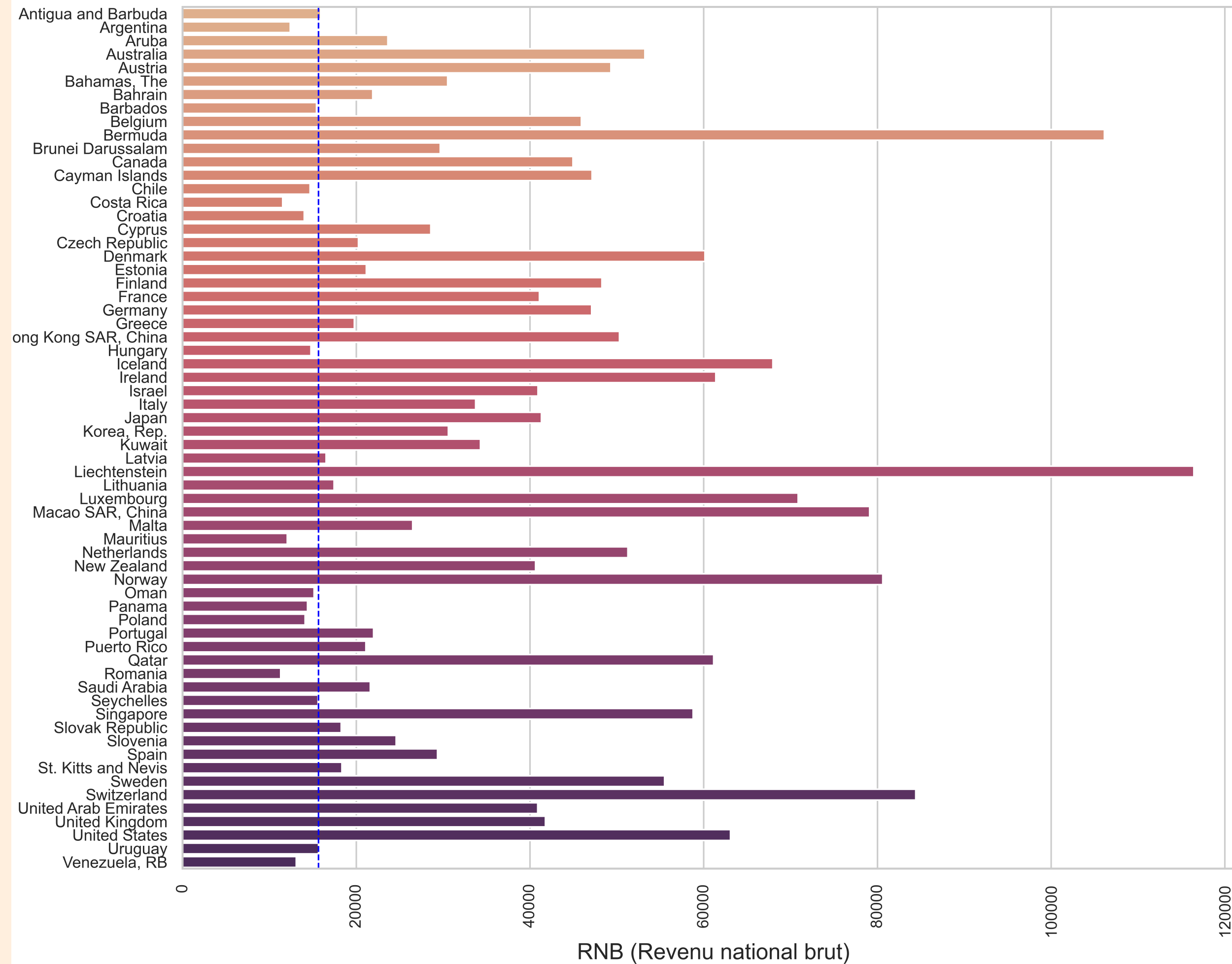
Analyse Exploratoire

RNB (Revenu national brut)



Le RNB en moyenne est de 15 644 \$. Il y a 51 pays dans la liste dont le RNB est plus élevé que la moyenne.

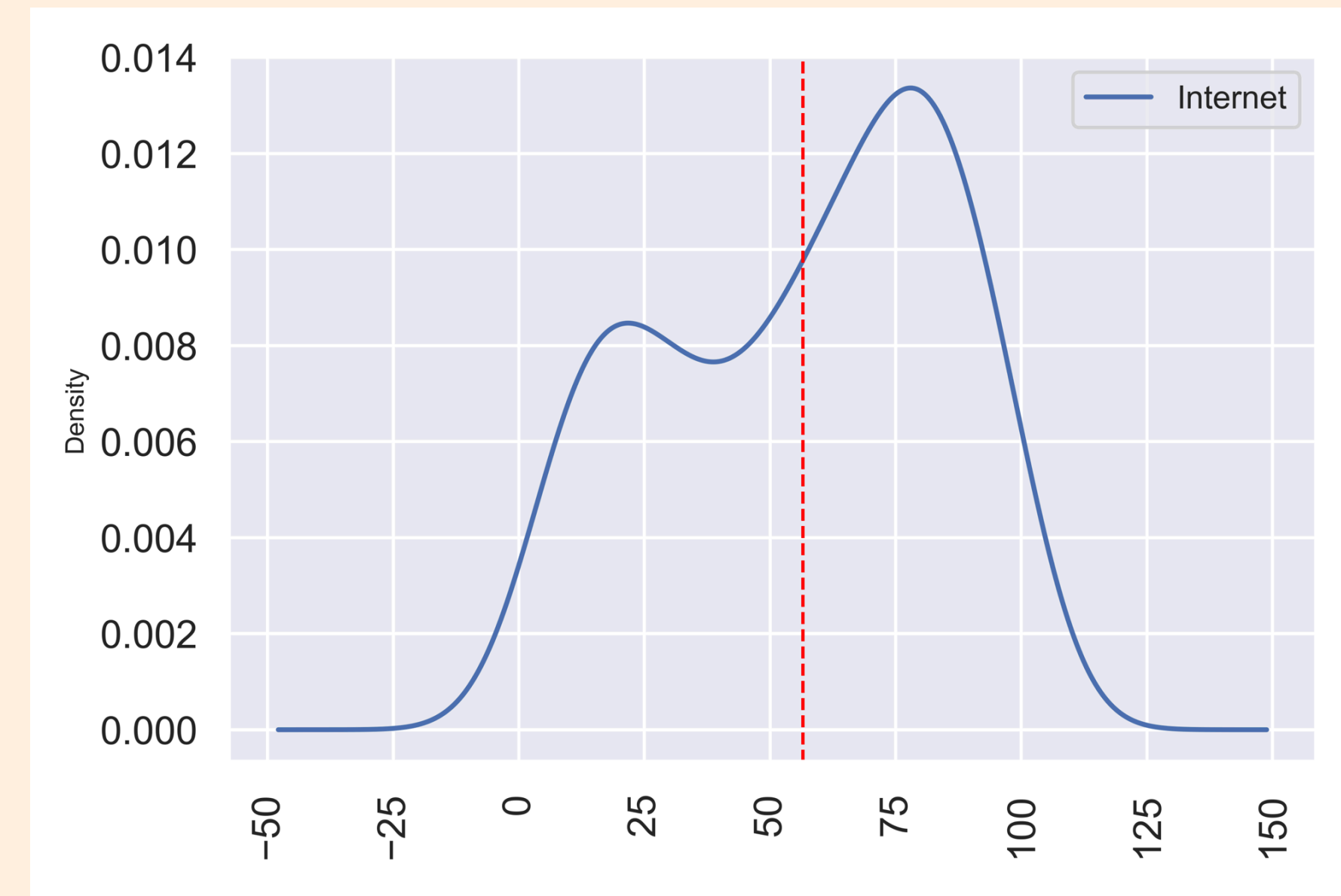
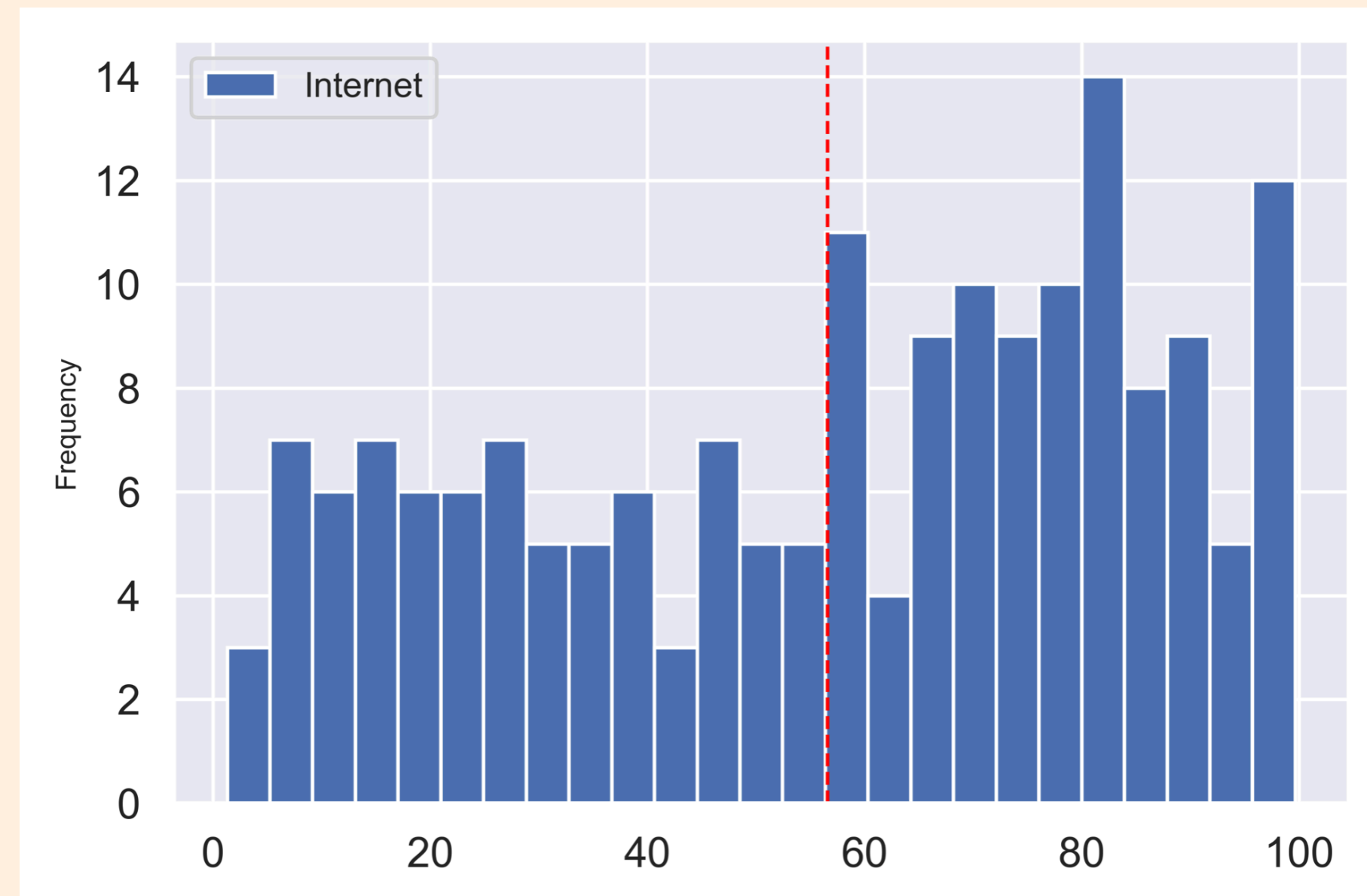
RNB (Revenu national brut)



Maintenant, nous apercevons un graphique représentant une partie des pays ayant un RNB supérieur à 11 000 dollars.

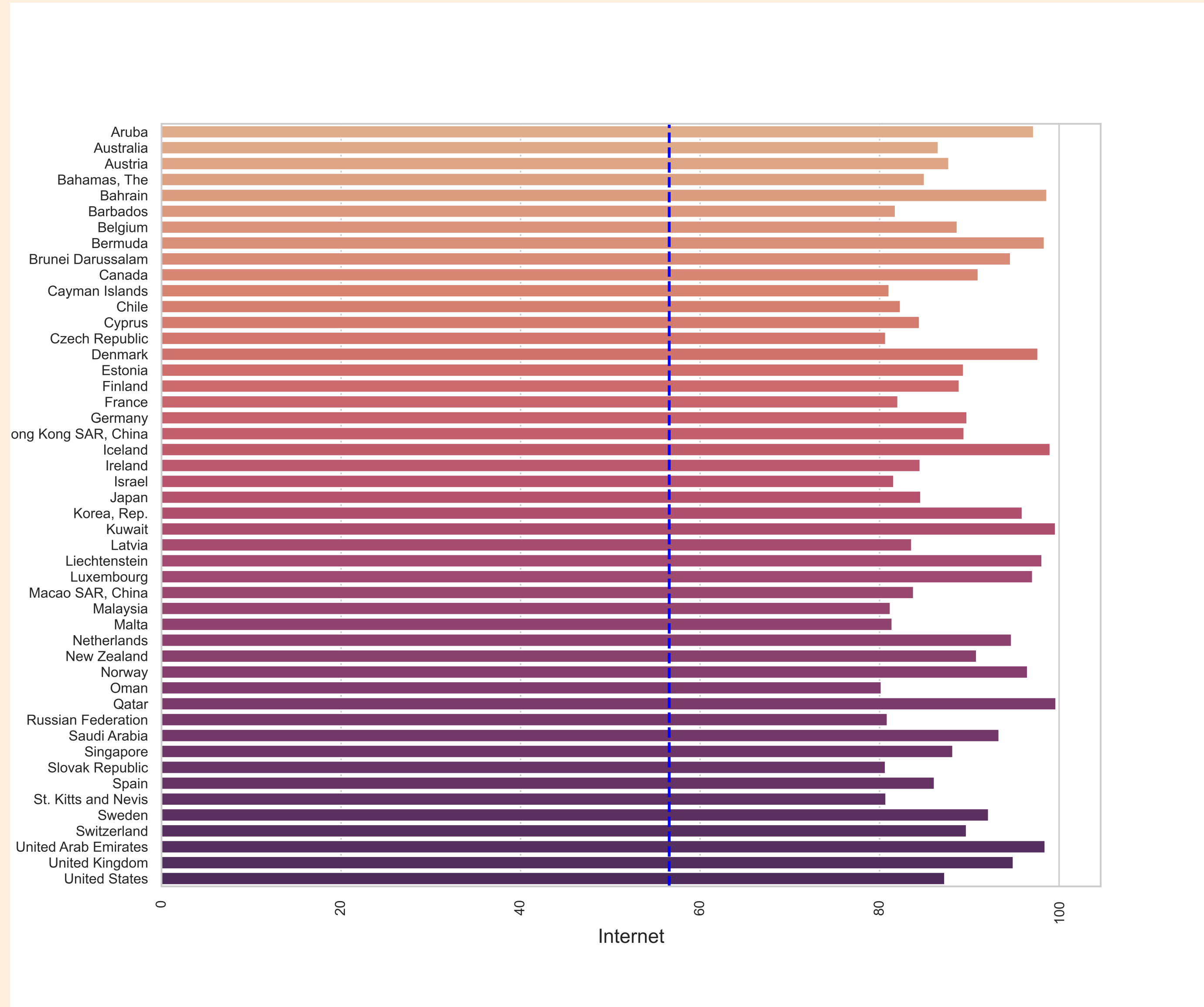
Par exemple, le RNB de la France et de l'Allemagne dépasse 40 000 dollars.

Internet (Le taux d'utilisation d'Internet)



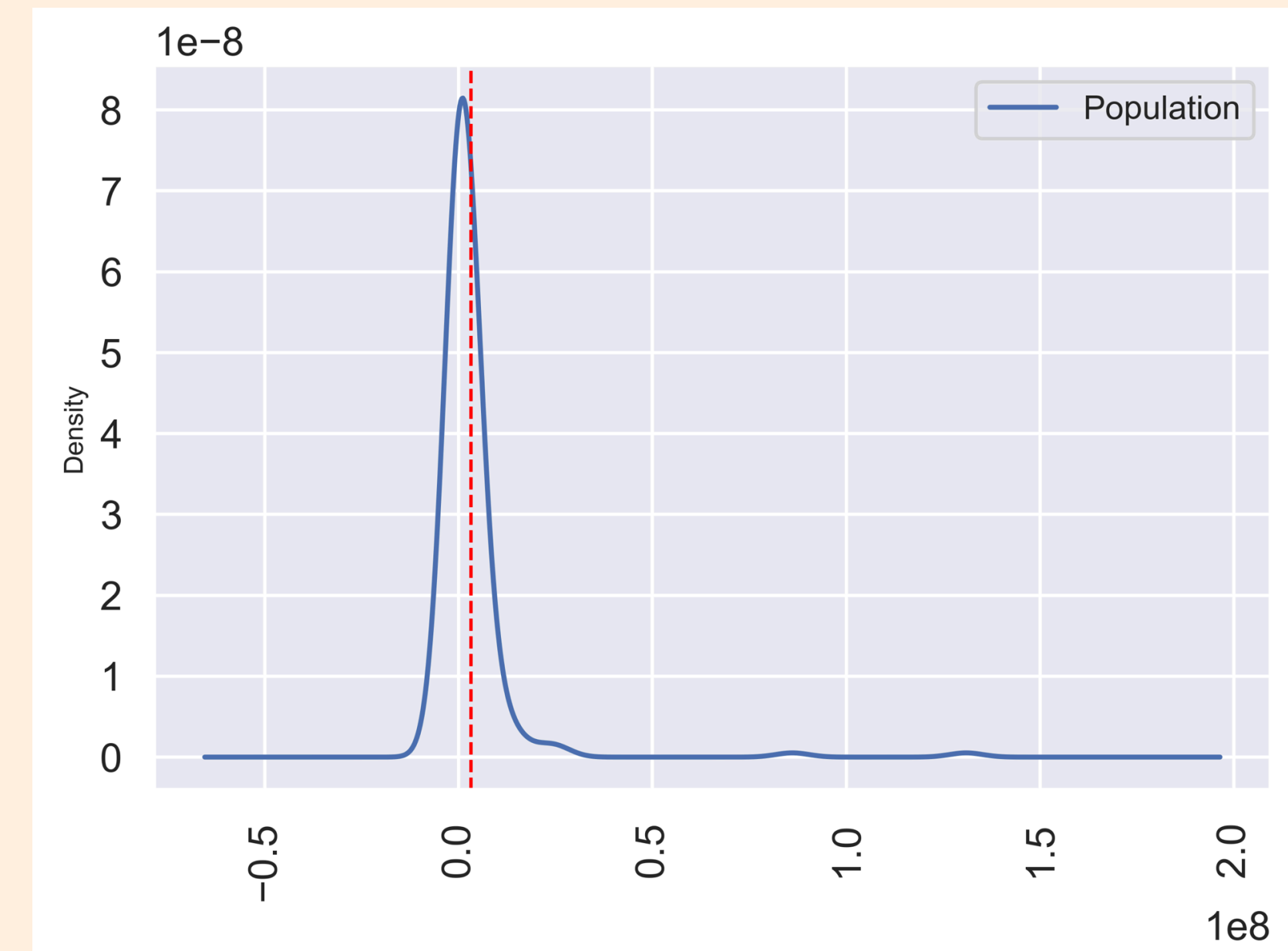
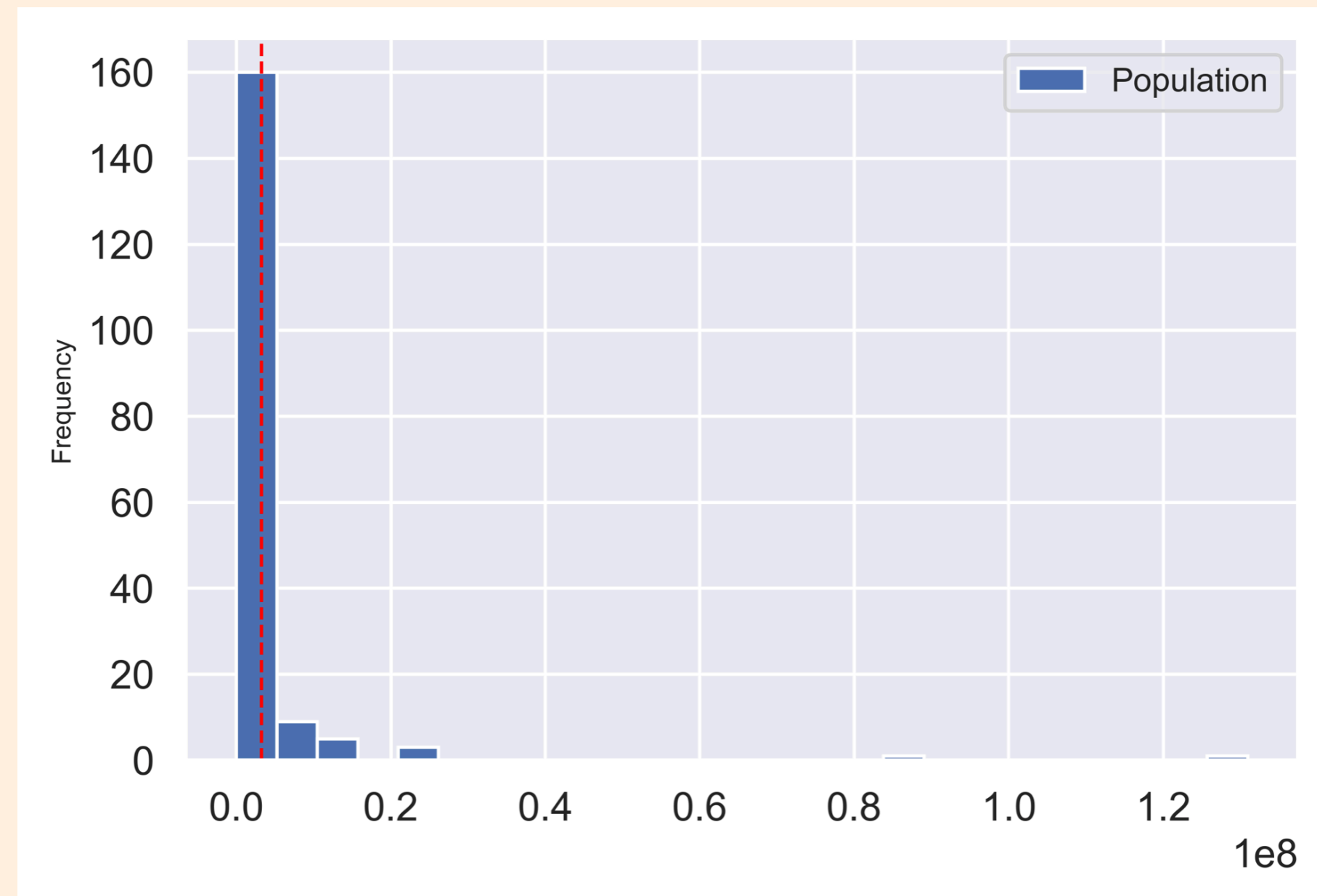
Le nombre d'utilisateurs d'Internet en moyenne est de 56 %. Il y a 101 pays dans la liste qui ont un taux plus élevé que 56 %.

Internet (Le taux d'utilisation d'Internet)



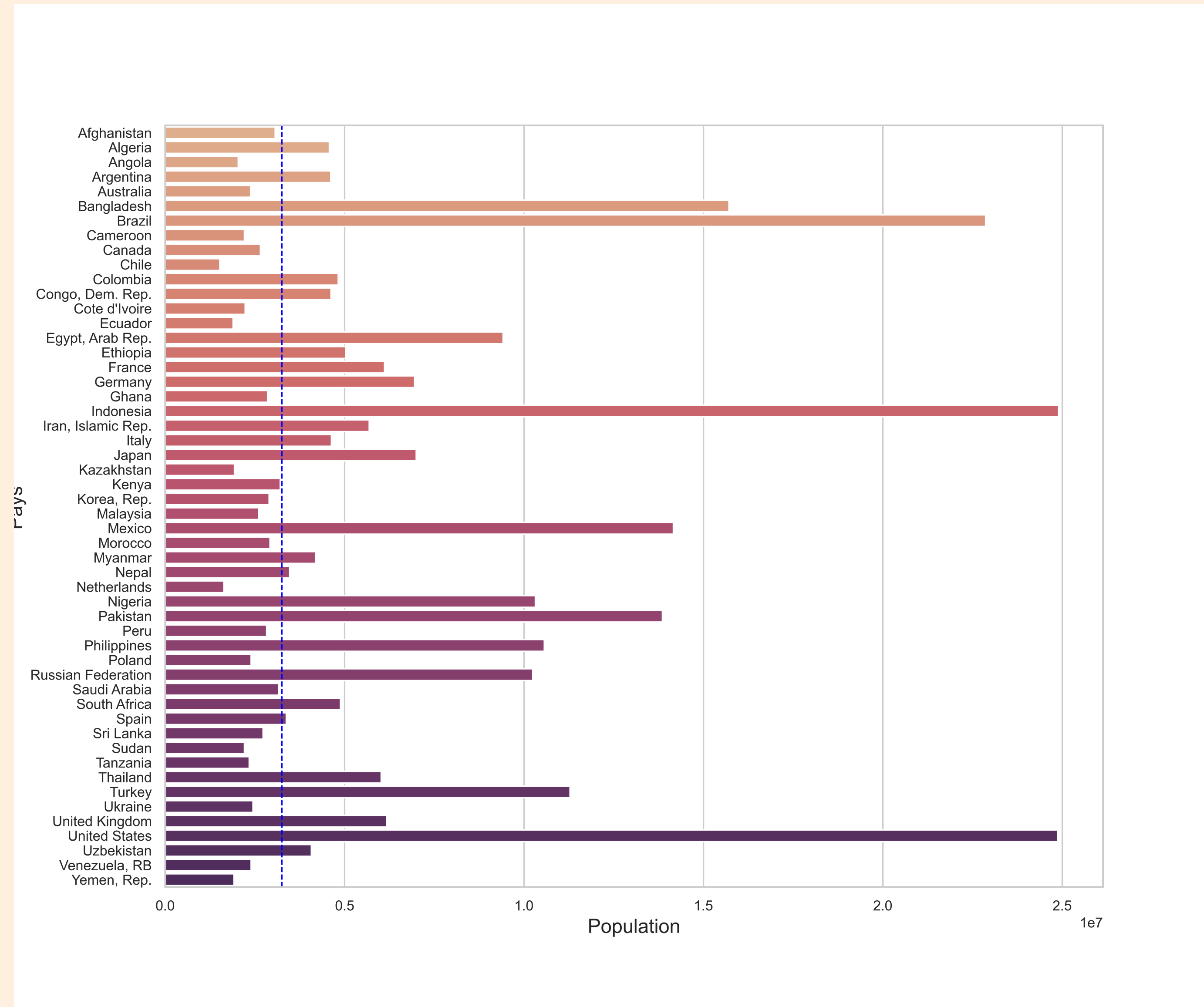
Le graphique des pays où le taux d'utilisation d'Internet est supérieur 80 %.

Population (Le nombre d'inscriptions dans l'enseignement secondaire)



La population moyenne est 3 248 108. Il y a 30 pays dans la liste dont la population dépasse la moyenne.

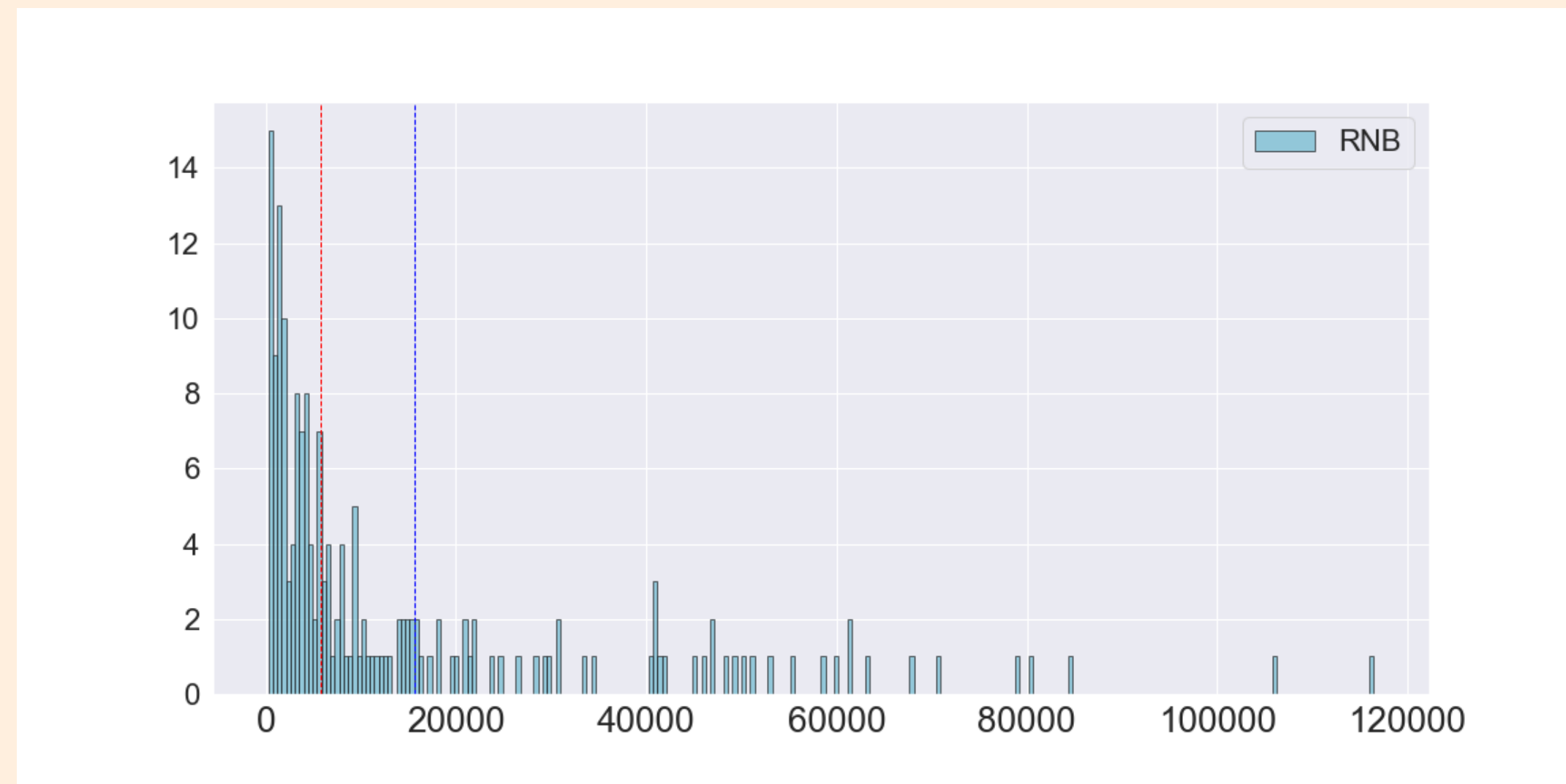
Population (Le nombre d'inscriptions dans l'enseignement secondaire)



Bien que la France et la Tanzanie comptent presque le même nombre d'habitants (environ 63 millions d'habitants), nous pouvons constater sur ce graphique que la France se situe au-dessus de la moyenne alors que Tanzanie se trouve en-dessous.

Nous pouvons en déduire qu'il y a des pays qui sont grands au regard de leur population entière, mais qui n'ont pas un système d'éducation élevé.

Méthode d'analyse



Pour éliminer les pays les moins intéressants au niveau des 3 indicateurs, c'est-à-dire,

- Le Revenu National Brut,
- Le taux d'utilisation d'Internet,
- La population d'étudiants de niveau lycée et université

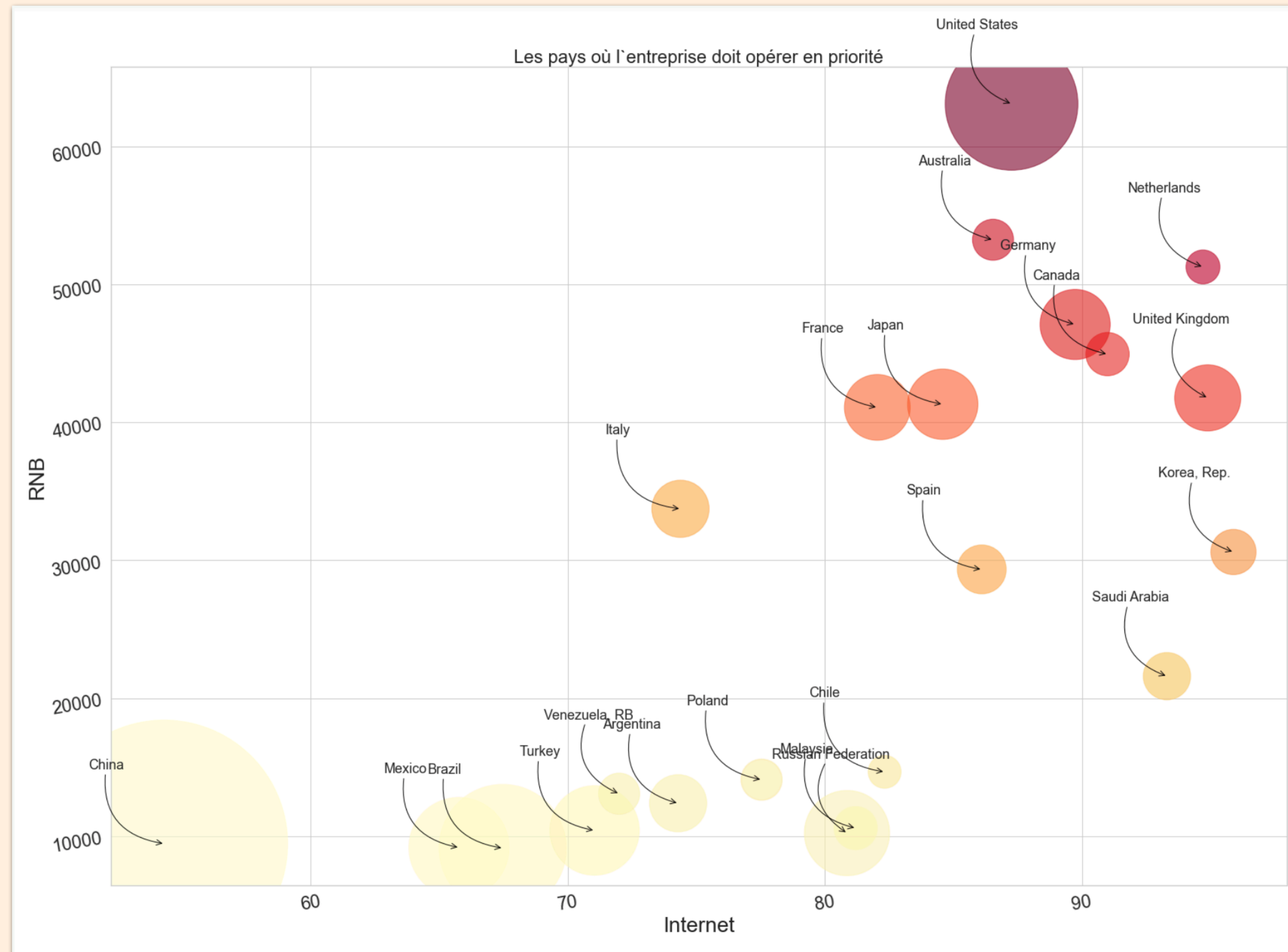
Nous prenons en compte la pré-analyse que nous avez faite sur les indicateurs.

Les pays avec un fort potentiel de clients sont précisés avec les conditions que nous avons déterminées en regardant la distribution des pays. Celles-ci sont de 9 000 \$ pour le RNB, de 50 % pour le taux d'utilisation d'internet, et pour la population d'étudiants de 15 millions. Ces contraintes se trouvent entre la médiane (ligne rouge) et la moyenne (ligne bleue).

Par exemple sur ce graphique, vous voyez la distribution du RNB avec la contrainte de 9 000 \$.

Les pays où l'entreprise doit opérer en priorité

Les pays où l'entreprise doit opérer en priorité



Ce graphique reprend le résultat de l'analyse des pays où il faut opérer en priorité.

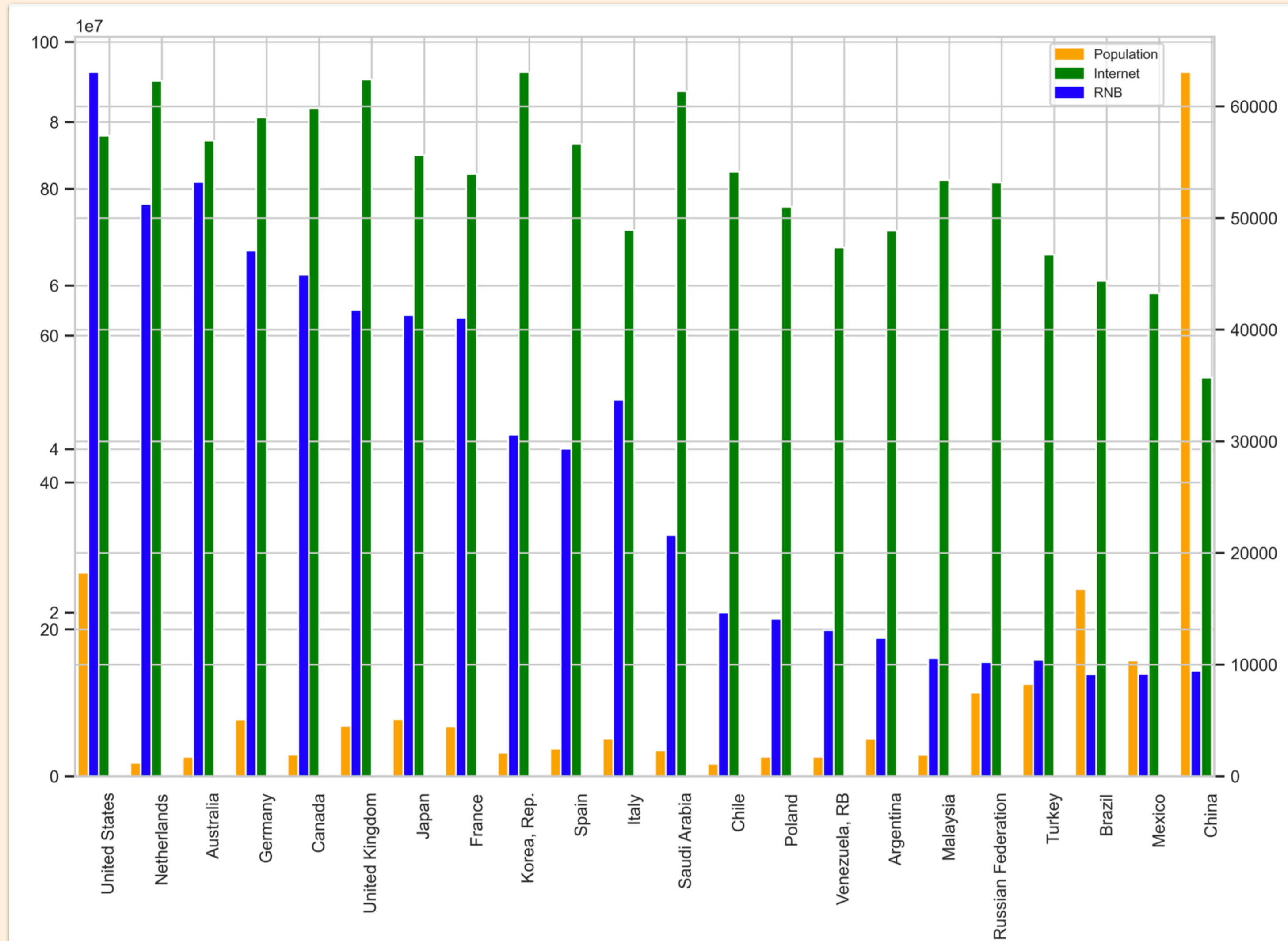
Notre liste contient 22 pays comprenant notamment les états-unis, l'Italie et la Chine.

Plus le pays est situé en haut du graphique, plus son RNB est élevé.

Plus le pays est situé à droite du graphique, plus son taux d'utilisation d'internet est élevé.

La taille des cercles est proportionnelle au nombre d'étudiants.

Les pays où l'entreprise doit opérer en priorité



Sur ce graphique, nous avons repris la liste des 22 pays.

Pour chaque pays, il y a trois barres :
une barre bleue représentant le RNB,
une barre verte représentant le taux d'utilisation
d'internet,
et une barre jaune représentant la population étudiante.

Par exemple, nous pouvons voir que la Chine a une
population très élevée pour un RNB un peu faible par
rapport aux autres pays de la liste.

A l'inverse, les états-unis ont une population faible par
rapport à la Chine mais un taux d'utilisation d'internet
très élevé.

Les pays où l'entreprise doit opérer en priorité

| | Population | RNB | Internet | Rank |
|-----------------------|--------------|---------|-----------|----------|
| Pays | | | | |
| United States | 2.487136e+07 | 63080.0 | 87.266113 | 5.504746 |
| Netherlands | 1.632486e+06 | 51260.0 | 94.712074 | 4.854941 |
| Australia | 2.379962e+06 | 53230.0 | 86.545049 | 4.606793 |
| Germany | 6.948797e+06 | 47090.0 | 89.739121 | 4.225815 |
| Canada | 2.654199e+06 | 44940.0 | 91.000000 | 4.089540 |
| United Kingdom | 6.173691e+06 | 41770.0 | 94.896742 | 3.963837 |
| Japan | 6.995597e+06 | 41310.0 | 84.588066 | 3.494333 |
| France | 6.109604e+06 | 41080.0 | 82.043534 | 3.370348 |
| Korea, Rep. | 2.893979e+06 | 30600.0 | 95.898292 | 2.934488 |
| Spain | 3.371069e+06 | 29340.0 | 86.107236 | 2.526386 |
| Italy | 4.630438e+06 | 33730.0 | 74.387183 | 2.509080 |

| | Population | RNB | Internet | Rank |
|---------------------------|------------|---------|-----------|----------|
| Pays | | | | |
| Saudi Arabia | 3159358.0 | 21600.0 | 93.310002 | 2.015496 |
| Chile | 1522145.0 | 14670.0 | 82.327487 | 1.207744 |
| Poland | 2392353.0 | 14100.0 | 77.541735 | 1.093338 |
| Venezuela, RB | 2391174.0 | 13080.0 | 72.000000 | 0.941760 |
| Argentina | 4612663.0 | 12390.0 | 74.294907 | 0.920514 |
| Malaysia | 2602216.0 | 10590.0 | 81.201707 | 0.859926 |
| Russian Federation | 10242404.0 | 10230.0 | 80.864722 | 0.827246 |
| Turkey | 11279561.0 | 10420.0 | 71.042761 | 0.740266 |
| Brazil | 22863908.0 | 9140.0 | 67.471285 | 0.616688 |
| Mexico | 14160635.0 | 9180.0 | 65.772634 | 0.603793 |
| China | 86101697.0 | 9460.0 | 54.300000 | 0.513678 |

La liste finale des pays où l'entreprise doit opérer en priorité.

Merci de votre attention