

FACULTÉ DES SCIENCES ET INGÉNIERIE

SORBONNE UNIVERSITÉ

---

# Rapport Business Intelligence

---

Emmanuel JUDITH

Wenshan WU

14 avril 2019

# Table des matières

1	Contexte . . . . .	2
2	Données . . . . .	2
3	Modélisation . . . . .	3
4	Visualisation et interprétation . . . . .	5
4.1	Population . . . . .	5
4.2	Espérance de vie . . . . .	7
4.3	Taux de fertilité . . . . .	9
4.4	Produit intérieur brut . . . . .	11
5	Conclusion . . . . .	12

# 1 Contexte

Les personnes âgées représentent une part importante de notre société. Nombre d'entre elles restent dans la vie active malgré leur grand âge, et après avoir travaillé elles souhaitent se reposer. Cependant toutes n'ont pas leur famille à disposition pour s'occuper d'elles ce qui a pour conséquences que nombre d'entre elles ont recours à une maison de retraite. Cela peut aussi être dû au peu de moyen pour payer un loyer ou encore le manque d'autonomie pour effectuer des tâches quotidiennes. Notre problématique est donc de savoir ou implémenter des maisons de retraite sur le continent européen en prenant en compte un certain nombre de paramètres. En effet, s'inscrire en maison de retraite requiert un coût il faut donc trouver un pays plutôt développé dans lequel les habitants possèdent les moyens de s'offrir ce service. De plus nous devons miser sur les pays à espérance de vie forte car les personnes encore jeunes ont potentiellement moins d'intérêt à s'y inscrire. Ce sont des exemples de questions que nous nous sommes posées en commençant ce projet, c'est pourquoi nous avons choisi des données permettant d'orienter notre réflexion dans ces directions.

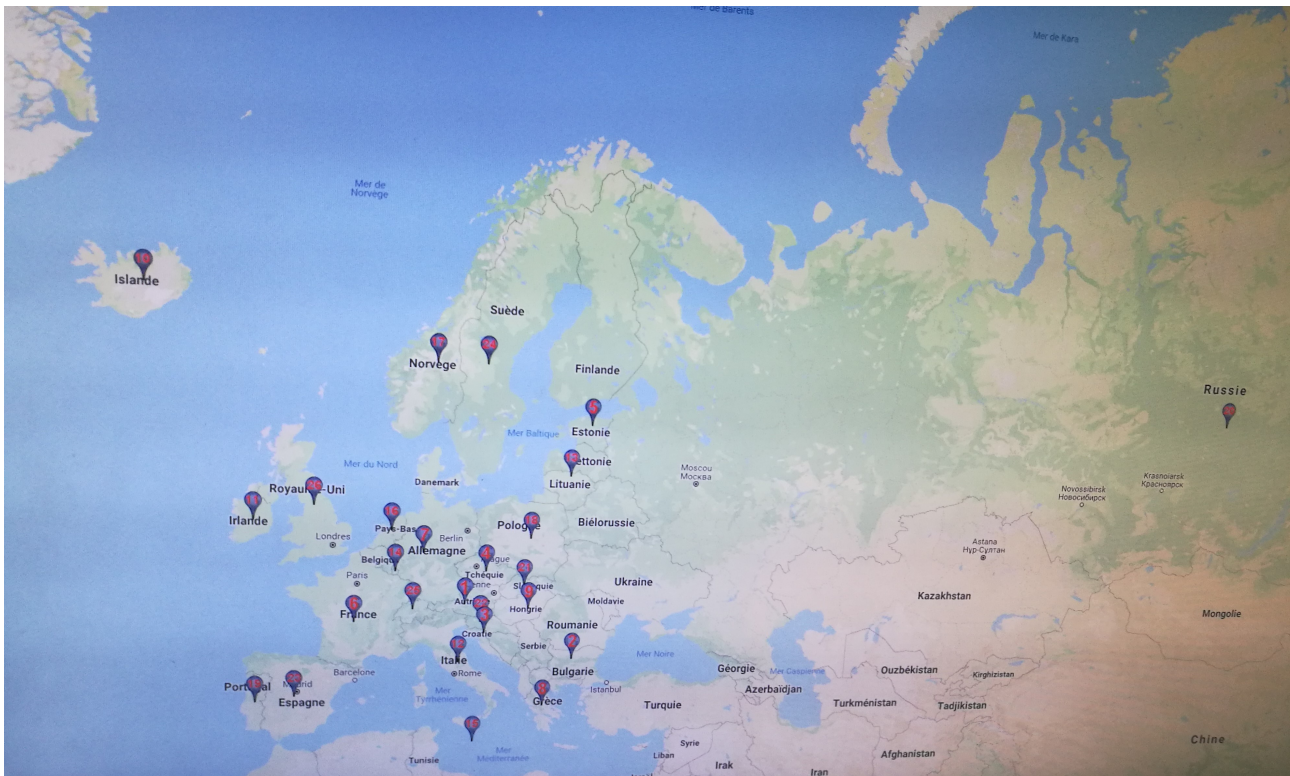
## 2 Données

Nous nous sommes appuyés sur plusieurs données pour faire notre analyse que nous allons vous présenter ainsi que le processus d'extraction. Tout d'abord *world.sql*, la base de données qui nous a été fourni et qui contient déjà beaucoup d'information sur un grand ensemble de pays du monde, tels que le nom et leur code.

Les données suivantes sont celles que nous avons trouvées et qui concernent tous les pays mondiaux sur les années 1960 à 2016. Nous avons uniquement extrait les 10 dernières années pour garder de la pertinence dans l'analyse et fait un filtrage pour ne garder que certains pays européens.

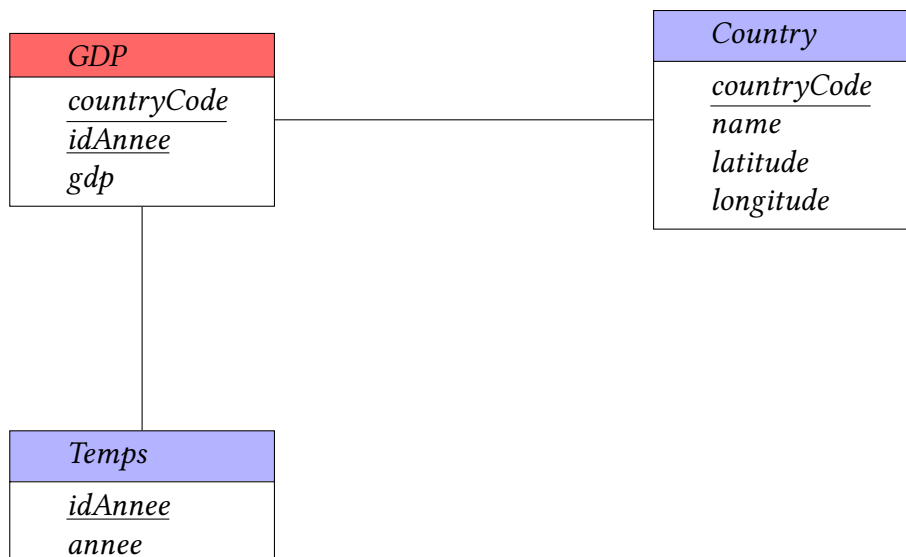
- *dp\_live.csv* est un fichier concernant le PIB de chaque pays selon deux mesures différentes. Nous avons gardé celle en mln\_usd car elle est la plus utilisée. Cette donnée est importante pour choisir les pays candidats dans lesquels implémenter notre projet.
- *life\_expectancy.csv* nous donne l'espérance de vie de la population. Nous devons choisir un pays où elle est assez élevée.
- *population.csv* contient le nombre d'habitants par pays. Cela nous permet d'avoir un aperçu global du nombre de personnes qui pourrait être touchées.
- *fertility\_rate.csv* montre le nombre d'enfant moyen par foyer. Ceci est un critère qui a également son importance.
- *average-latitude-longitude-countries.csv* pour avoir les coordonnées géographiques moyenne des pays.

Nous avons fait nos analyses sur le pays indiqués sur la carte ci-dessous.



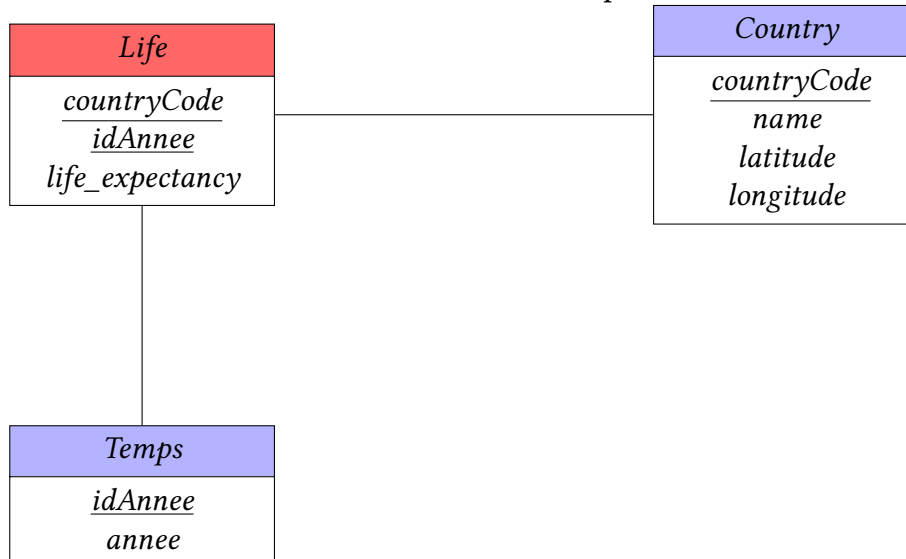
### 3 Modélisation

Lors de l'extraction de nos données nous avons définis les faits et les dimensions en rapport avec notre problématique. Quatre faits sont donc ressortis : l'espérance de vie, la valeur du PIB, le nombre d'habitants et le taux de fertilité. Les dimensions sont au nombre de deux, le pays et l'année.



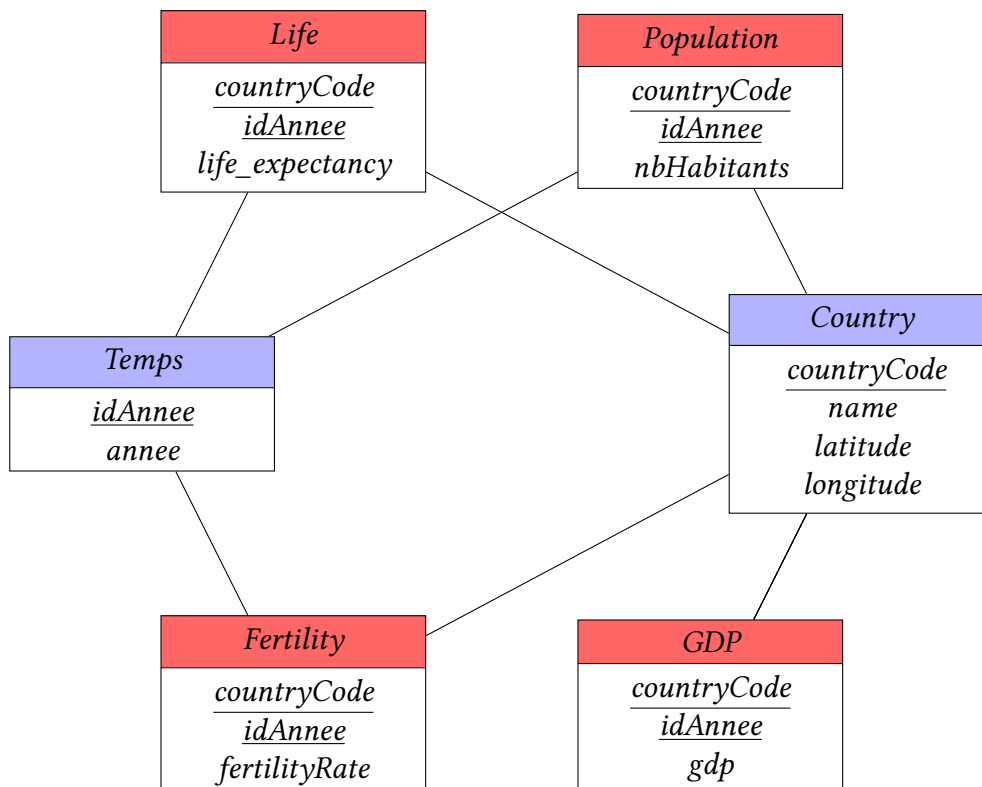
Voici le schéma en étoile correspondant au fait *GDP*, qui représente le sujet d'analyse qu'est le PIB. Les dimensions sont en bleu et le fait en rouge. Nous voyons bien que le PIB sera considéré en fonction d'une année et d'un pays, tous deux ayant leur identifiant présent dans sa table. L'attribut *gdp* symbolise la valeur du PIB. Dans la table *Country* nous avons *countryCode* le code du pays en format ISO (trois lettres) comme clé primaire, *name* le nom du pays en anglais tel qu'il était noté dans la base, ainsi que la latitude et la longitude pour pouvoir l'identifier sur une carte. En ce qui concerne la table *Temps*, nous avons dû fabriquer un identifiant pour chaque année en utilisant

plusieurs fonctionnalités du logiciel Pentaho. D'abord transposer les années qui étaient en colonnes pour les mettre en ligne puis leur attribuer un identifiant unique, sachant qu'une année se répétait autant de fois qu'il y avait de pays puisque dans les différents data sets l'analyse se faisait sur période s'étalant sur une trentaine d'années. Il a fallu procéder de la même façon pour tous les faits.



Ceci est le schéma en étoile pour le fait *Life*. Il est semblable au schéma précédent, seul l'attribut *life\_expectancy*, l'espérance de vie en années, est différent.

Les autres schémas étant similaires il est plus pertinent de montrer tous les faits en même temps sur un schéma en constellation :



Après extraction des données dans notre datawarehouse nous obtenons la représentation en logique dénormalisée ci-contre :

Life(countryCode,idAnnee,life\_expectancy)

Population(countryCode,idAnnee,nbHabitants)

Fertility(countryCode,idAnnee,fertilityRate)

GDP(countryCode,idAnnee,gdp)

Country(countryCode,name,latitude,longitude)

Temps(idAnnee,annee)

## 4 Visualisation et interprétation

Suite à l'extraction de nos données, nous avons pu obtenir plusieurs résultats sous la forme de graphes en utilisant le logiciel *Excel*. Nous allons par la suite vous les présenter un à un, et vous faire part des informations tirées ainsi que des corrélations déduites afin de répondre à notre problématique.

### 4.1 Population

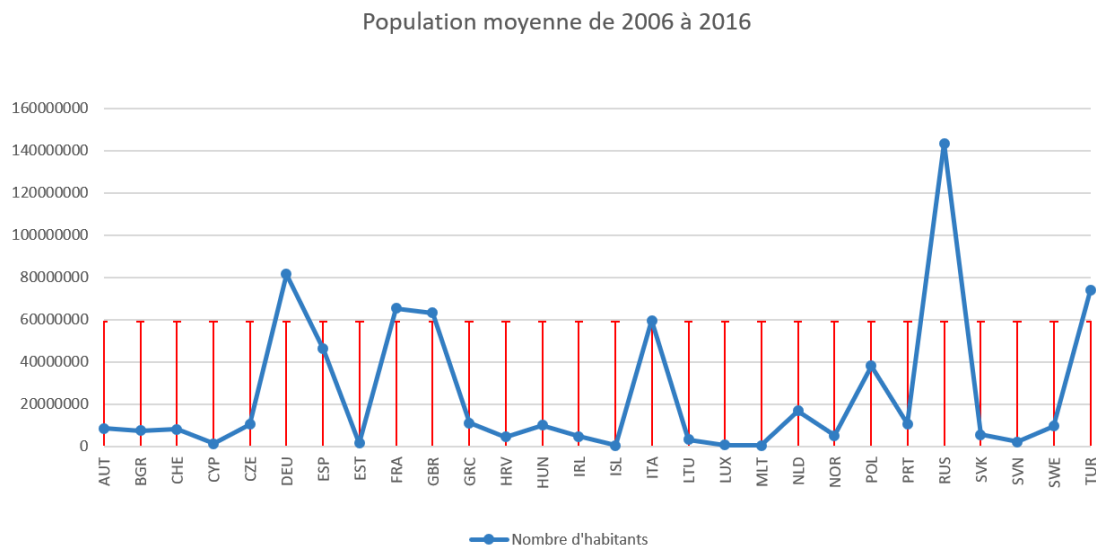


FIGURE 1 – Population moyenne de 2006 à 2016 en UE

Cette figure 1 représente la population moyenne de la majorité des pays de l'Union Européenne de 2006 à 2016 reflétée par la courbe bleu, tandis que la partie rouge représente l'écart-type de la moyenne afin de nous donner une idée de la valeur de la population moyenne. Nous observons dans un premier temps que certains pays se trouvent être très peuplés, en particulier l'Allemagne, la France ou encore la Russie avec respectivement environ 80M, 62M et plus de 140M d'habitants. Cela se remarque aisément au vu de leur moyennes qui se trouvent au-dessus de l'écart-type. Nous décidons donc de choisir ces pays comme candidats potentiels dans le cadre de notre problématique mais avant cela, nous avons analysé à chacun leur évolutions respectives au cours des dix dernières années à partir de 2016, voici les informations obtenues.

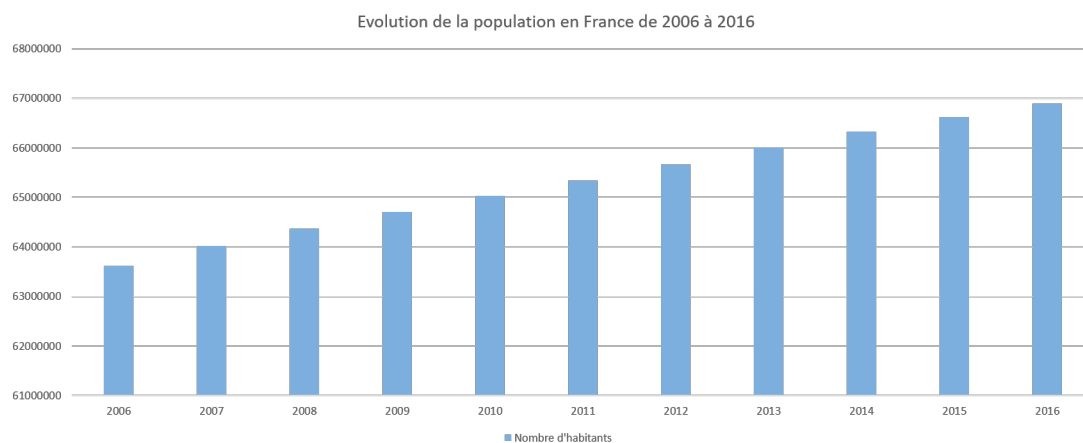


FIGURE 2 – Évolution de la population en France de 2006 à 2016

Cette figure 2 illustre l'évolution de la population en France depuis 2006 à 2016, on observe une évolution constante et progressive où en 2006, le nombre d'habitants était d'environ 63 600 000 et en 2016 elle a atteint un pic à quasiment 67M d'habitants. D'après ces résultats, cela nous conforte dans le choix de la France.

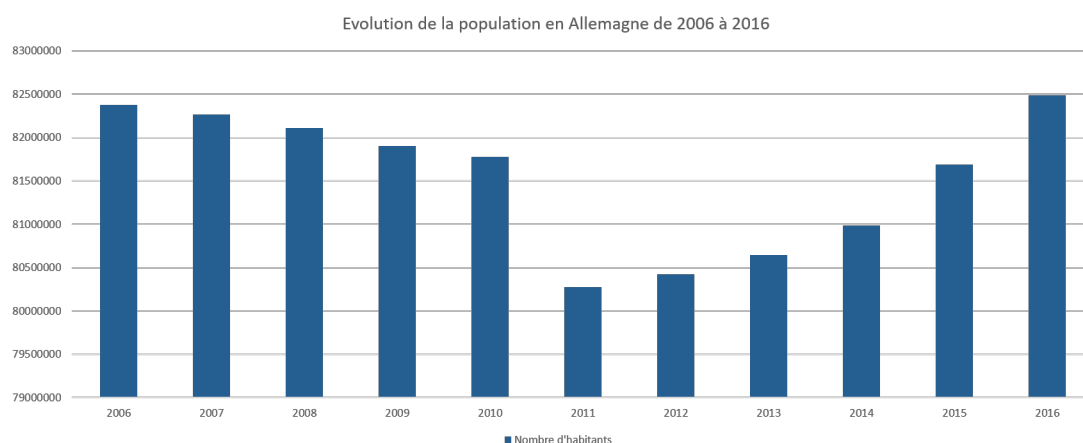


FIGURE 3 – Évolution de la population en Allemagne de 2006 à 2016

La figure 3 illustre l'évolution de la population en Allemagne depuis 2006 à 2016, on remarque un nombre d'habitants très élevé depuis 2006 avec plus de 82 300 000, malgré une très faible baisse en 2011 d'environ 1M d'habitants, elle reste néanmoins très peuplée avec quasiment 82 500 000 habitants en 2016 et donc un très bon candidat.

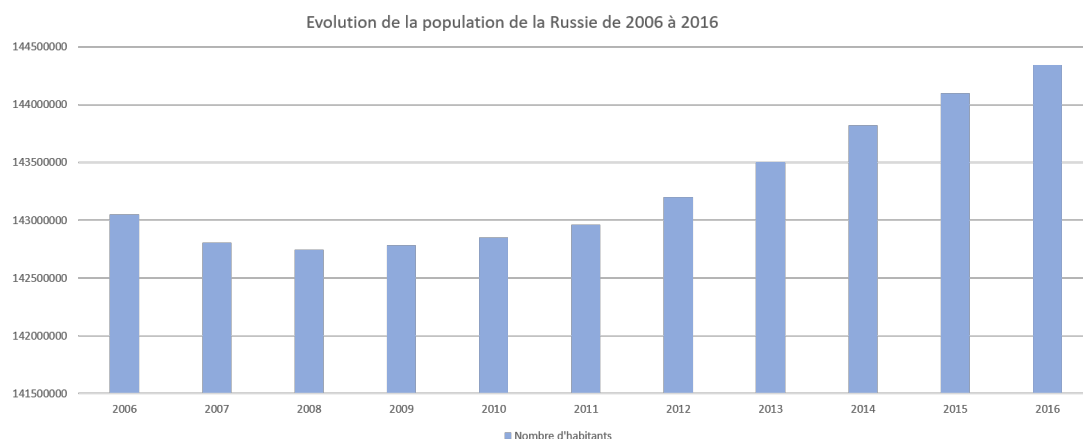


FIGURE 4 – Évolution de la population en Russie de 2006 à 2016

Cette figure 4 montre l'évolution de la population en Russie depuis 2006 à 2016, comme prévu elle détient le plus grand nombre d'habitants parmi tous nos pays avec plus de 143M en 2006 et 144M en 2016. Cet histogramme montre une évolution croissante au fil des années et donc en fait un candidat potentiel.

## 4.2 Espérance de vie

Intéressons nous maintenant à un deuxième paramètre qui se trouve être l'espérance de vie d'un pays. Ce dernier est en lien direct avec le nombre d'habitants que nous avons étudié et comme précédemment, nous allons vous présenter nos résultats à travers la même forme d'étude.

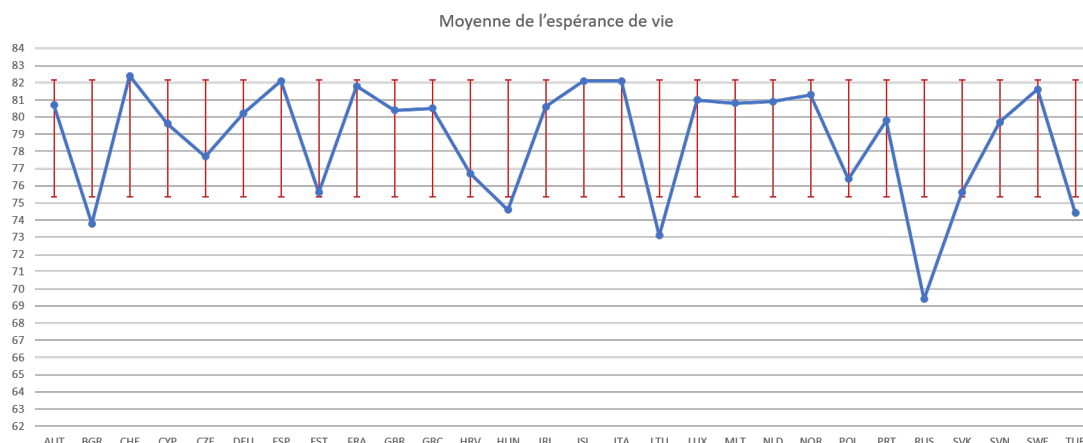


FIGURE 5 – Espérance de vie moyenne de 2006 à 2016 en UE

Cette figure 5 nous présente l'espérance de vie moyenne sur l'ensemble des pays de l'Union Européenne de 2006 à 2016, la courbe bleu définit pour chaque pays la valeur moyenne et la partie rouge l'écart-type de la moyenne. On observe dans un premier temps que la plupart des pays se trouvent dans l'intervalle de l'écart-type révélant une bonne homogénéité de l'espérance de vie. On remarque ensuite que certains pays comme la Suisse, France ou encore l'Italie montrent une espérance plus élevée que la moyenne, supérieur ou égale à 82 ans. Et pour finir, une observation flagrante a pu être réalisée en observant l'espérance de la Russie, cette dernière se trouve être très en-dessous de la moyenne avec moins de 70 ans par habitant. De ce fait, nous pouvons exclure la Russie comme



candidat potentiel étant donné sa très faible espérance de vie qui constitue un atout majeur à la résolution de notre problématique. Par contre, nous pouvons nous intéresser à la Suisse.

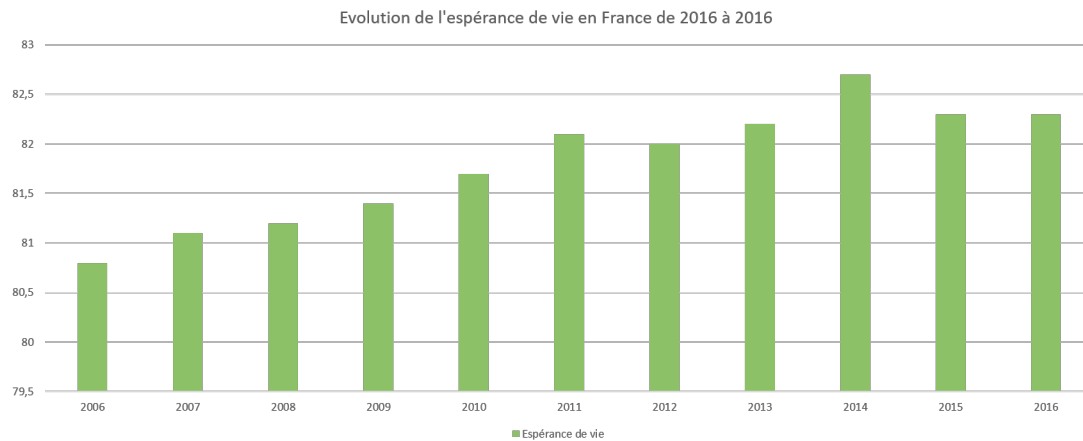


FIGURE 6 – Évolution de l'espérance de vie en France de 2006 à 2016

Cette figure 6 illustre l'évolution de l'espérance de vie en France depuis 2006 à 2016, on observe une évolution croissante depuis 2006 jusqu'à 2016 avec un pic de 82.7 ans en 2014. On en déduit que cette hausse de l'espérance de vie est en corrélation directe avec celle de la population par une évolution croissante des deux facteurs donc la France reste un bon candidat.

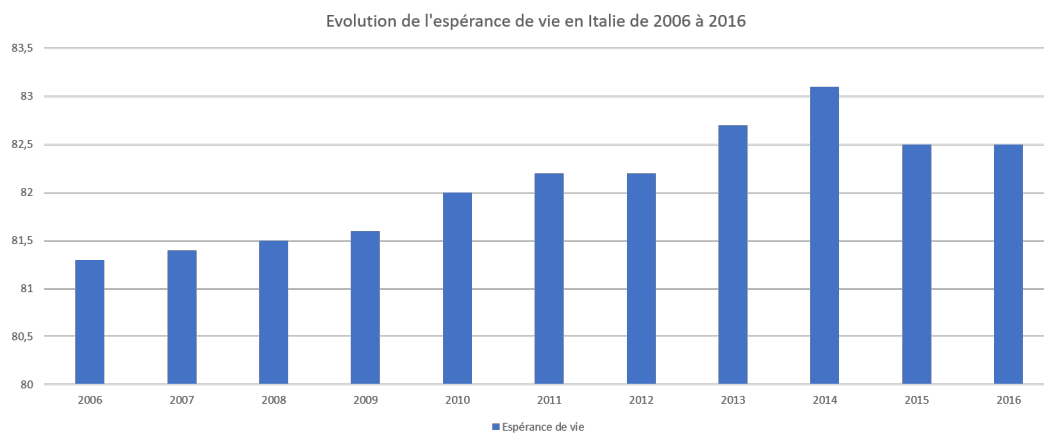


FIGURE 7 – Évolution de l'espérance de vie en Italie de 2006 à 2016

La figure 7 montre l'évolution de l'espérance de vie en Italie de 2006 à 2016, on observe une hausse depuis 2006 à 2016 avec 81.3 ans en 2006 et 82.5 ans en 2016. Ces chiffres montrent donc une évolution significative et nous permettent de considérer l'Italie comme un pays potentiel.

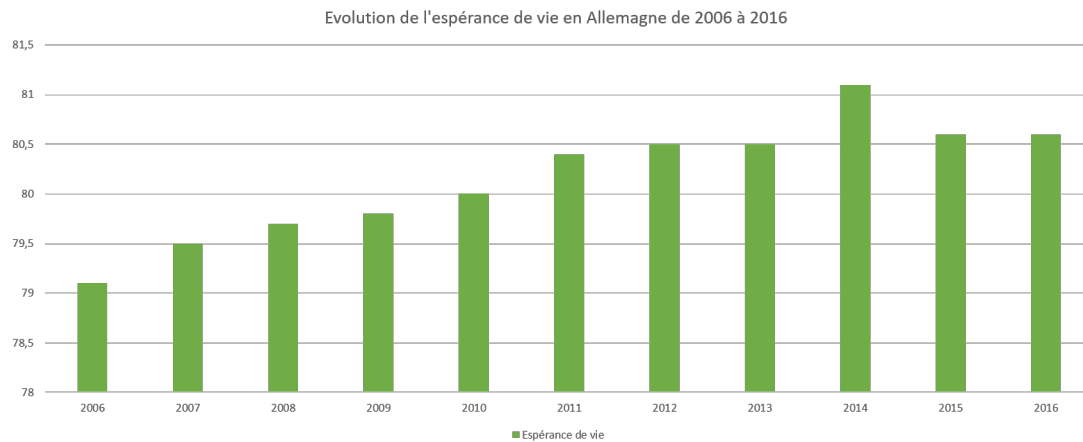


FIGURE 8 – Évolution de l'espérance de vie en Allemagne de 2006 à 2016

Cette figure 8 illustre l'évolution de l'espérance de vie en Allemagne depuis 2006 à 2016, comme pour la France on observe une évolution semblable, c'est à dire continue et croissante. Ainsi l'Allemagne reste un bon candidat.

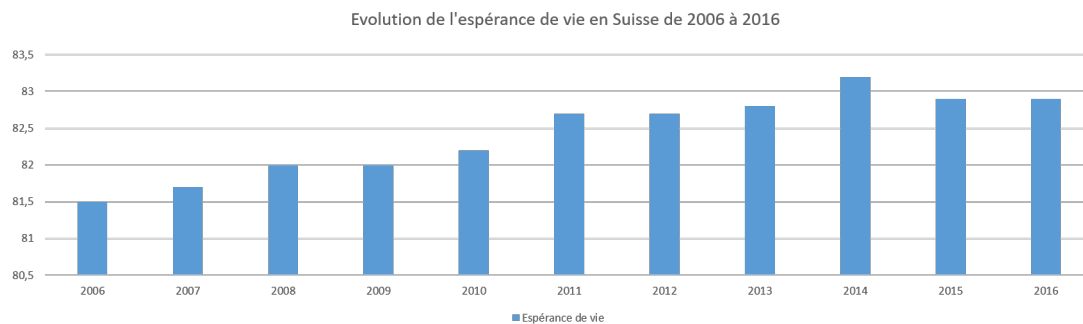


FIGURE 9 – Évolution de l'espérance de vie en Suisse de 2006 à 2016

La figure 9 montre l'évolution de l'espérance de vie de la Suisse depuis 2006 à 2016, on applique le même raisonnement que pour les pays précédents sur l'évolution de l'espérance. On décide ici de prendre en compte la Suisse dans la suite de notre interprétation du fait de son fort taux d'espérance malgré un nombre d'habitants moins élevé que les autres pays candidats comme nous avons pu le remarquer dans la figure 1.

### 4.3 Taux de fertilité

Notre troisième paramètre concernera le taux de fertilité, c'est à dire le nombre d'enfants par foyer, ce paramètre est également très important dans le cadre de notre évaluation, il se trouve être en lien direct avec le nombre d'habitants d'un pays ainsi que son espérance de vie. De nouveau, la même approche sera faite.

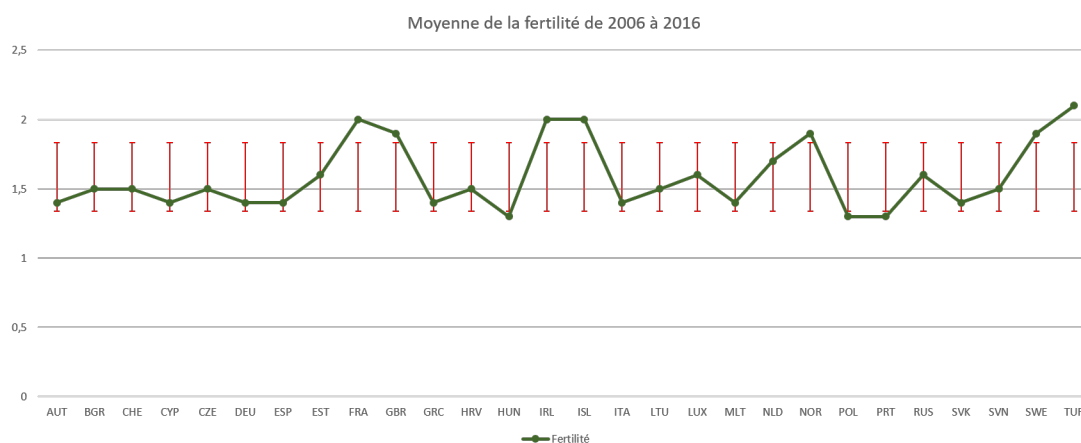


FIGURE 10 – Moyenne de la fertilité dans les pays de l’UE de 2006 à 2016

La figure 10 nous montre la moyenne de la fertilité dans l’ensemble des pays de l’UE de 2006 à 2016, la courbe verte représente le nombre moyen d’enfants par pays et la partie rouge représente l’écart-type de la moyenne. On observe dans un premier temps que la répartition du taux de fertilité est homogène, l’ensemble des valeurs sont comprises dans l’écart-type avec certaines qui atteignent des valeurs supérieures comme la France, l’Irlande ou bien l’Islande qui sont à 2 enfants en moyenne. Contrairement à la figure 5, il n’y a pas de valeurs aberrantes comme nous l’avons constaté pour la Russie, on n’élimine donc aucun de nos candidats et décidons d’établir leur histogrammes afin d’étudier l’évolution de leur taux de fertilité durant ces dix dernières années depuis 2006.

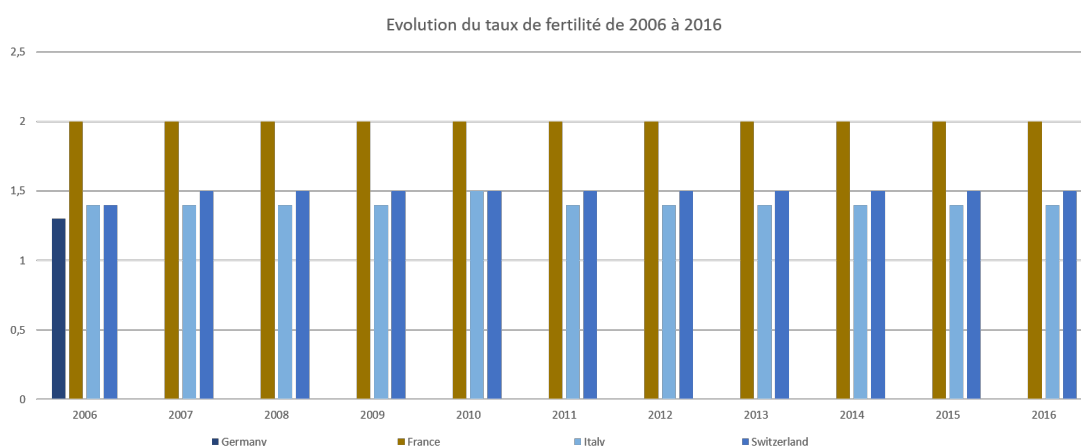


FIGURE 11 – Évolution du taux de fertilité en France, Allemagne, Suisse et Italie de 2006 à 2016

La figure 11 présente un histogramme groupé de plusieurs pays sur l’évolution du taux de fertilité de 2006 à 2016, on observe une forte régularité sur les dix années pour tous les pays, par exemple la France est constamment à 2 enfants en moyenne ou bien la Suisse à 1.5. Ces résultats consolident notre avancé par le fait qu’ils ne révèlent pas d’anomalie ou contradiction. Au vu de notre avancement actuel, nous sommes en mesure de faire la corrélation des trois derniers paramètres évalués qui sont la population, l’espérance de vie ainsi que le taux de fertilité. En effet, les graphes présentés montrent une hausse constante des trois paramètres sur plusieurs pays et donc de pouvoir considérer ces pays comme candidats potentiels.

## 4.4 Produit intérieur brut

Pour notre dernière évaluation, nous allons considérer le produit intérieur brut, aussi connu sous le nom de PIB. Ce critère d'évaluation est également très important étant donné qu'il est en lien avec ceux étudiés précédemment et qu'il nous permettra de consolider notre étude.

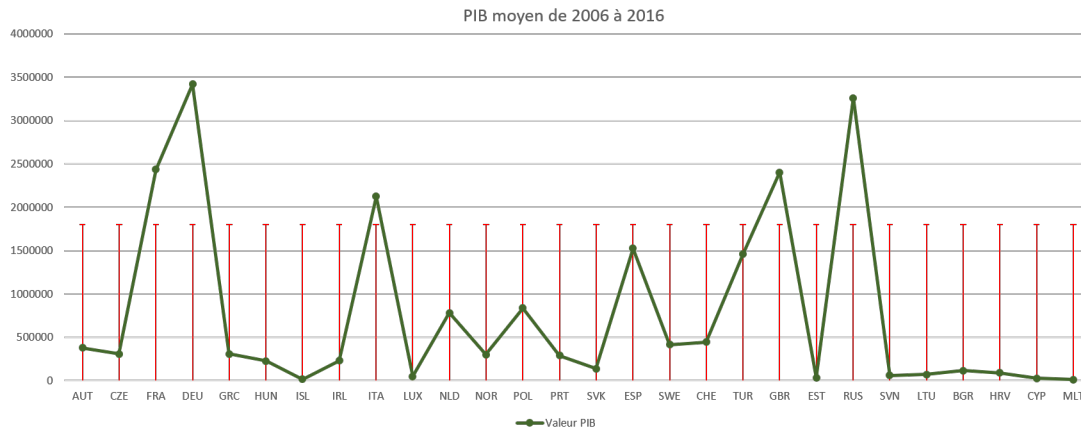


FIGURE 12 – PIB moyen des pays de l'UE de 2006 à 2016

La figure 12 illustre le PIB moyen de l'ensemble des pays de l'Union Européenne sur dix années de 2006 à 2016, la courbe verte représente la valeur moyenne du PIB pour chaque pays et la partie rouge représente l'écart-type de la moyenne. On observe une petite disparité avec les pays tels que la France, Allemagne ou encore Russie qui présentent un fort taux de PIB avec respectivement plus de 2.4, 3.4 et 3.2 milliards de dollars mais comme dit précédemment nous avons décidé d'exclure la Russie pour son sa faible espérance de vie et donc nous ne le comprendrons pas en compte dans notre analyse.

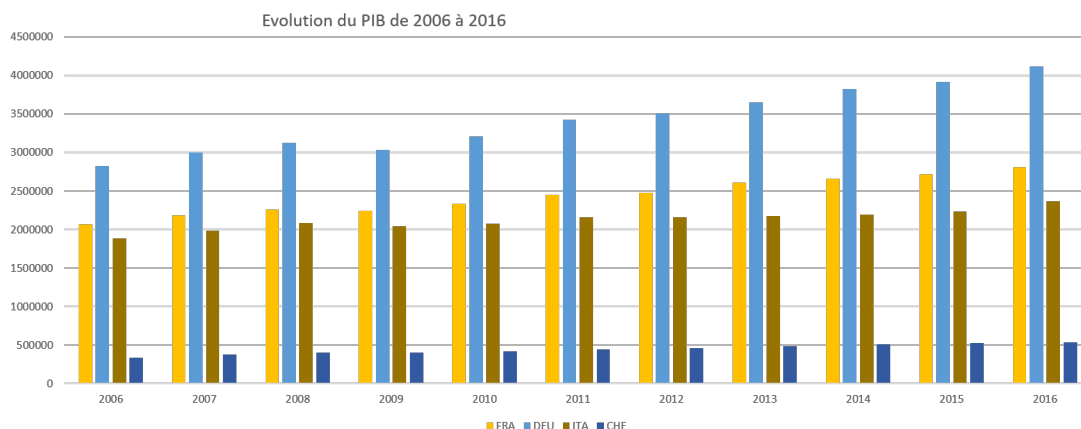


FIGURE 13 – Évolution du PIB de 2006 à 2016

Ce dernier graphique présente l'évolution du PIB de la France, Allemagne, Italie et Suisse de 2006 à 2016. On observe une évolution croissante et continue sur tous les pays présents, on remarque que le PIB de la Suisse est inférieur à celui des autres même s'il est constant sur les dix années.

## 5 Conclusion

Nous avons analysé dans un premier temps le nombre d'habitants des pays européens, suite à cela nous en avons sélectionné ceux qui sortaient du lot, c'est à dire qui présentaient un nombre d'habitants supérieur à la moyenne. Nos graphiques se sont ensuite centrés sur ces pays lors de l'évaluation des critères suivants, nous avons souvent remarqué une corrélation entre l'espérance de vie, le taux de fertilité ainsi que la population. Ce qui nous a conforté dans le choix de nos potentiels candidats dans le cadre de notre implémentation cependant nous nous sommes également aperçus que ce n'était pas toujours le cas avec la Russie qui présente un grand nombre d'habitants mais une espérance de vie bien inférieure à la moyenne. Pour finir, nous nous sommes attaqués à l'étude sur le PIB, nous n'avons décelé aucune forte incohérence sur les chiffres concernant nos candidats, ce qui nous amène à dire qu'il existe une forte corrélation entre ces quatre paramètres. Nous en avons conclu la chose suivante suite à nos interprétations, les pays tels que la France, Allemagne, l'Italie ou encore la Suisse se révèlent être de très bons candidats pour être sujet d'implémentation de maisons de retraite. Plus en général, nous souhaiterions privilégier les pays ouest occidentaux par rapport à ceux qui se situent à l'est oriental.

# Bibliographie

- [1] Data on Kaggle  
<https://www.kaggle.com/farazrahman/population-life-expectancy-fertility-trends/data>
- [2] Average latitude longitude countries  
<https://github.com/albertyw/avenews/blob/master/old/data/average-latitude-longitude-countries.csv>