

# Projet Data Warehouse

## Thématique - Analyse mondiale

Quels sont les facteurs expliquant les dynamiques de population à l'échelle mondiale ?

---

**Santhos Arichandra, Almehdi Krisni**

**Master 1 DAC - 2021/2022**

**Enseignants: Nicolas Castanet & Laure Soulier**

# Introduction

L'homme est devenu au fil des siècles un individu sédentaire. Il cherche à s'établir dans le milieu correspondant le plus à ses attentes et besoins sur différents points, tels que sa situation salariale ou l'espace culturel du milieu. Il arrive parfois qu'il se déplace, que ce soit dans la métropole la plus proche ou dans un pays de l'autre côté du monde afin de répondre à ses exigences.

Il y a donc un intérêt concret à comprendre les motivations derrière les déplacements à l'échelle mondiale, qu'il s'agisse de l'immigration ou de l'émigration, que ce soit dans le cadre d'une étude gouvernementale ou bien de la mise en place de business. Par exemple, si l'on remarque qu'une population précise s'expatrie fortement vers une destination, il serait envisageable d'implémenter des boutiques vendant des produits de leurs produits d'origine.

La thématique principale du projet est de réaliser une analyse à l'échelle mondiale, de comparer des métriques à l'échelle des continents ou des pays. On se concentre alors sur des métriques économiques, de développement humain et géopolitiques. Il y a donc un besoin de créer des dimensions sur lesquelles nous étudierons les différents.

Afin d'illustrer le projet et la recherche, il sera également nécessaire de générer des illustrations afin de faciliter la compréhension, qu'il s'agisse de tableaux, de cartes mondiales ou de dashboards. Pour cela, on utilise Dataiku, un logiciel permettant de créer, fournir et représenter des données. On y transfère les fichiers contenant les données, soit des fichiers au format **.xls** ou **.json**, afin de pouvoir les visualiser dans différentes dimensions, ou bien créer les illustrations dont nous avons besoin.

On s'intéresse alors dans un premier temps à la collecte des données. Il existe plusieurs interfaces en ligne permettant de récupérer des données à propos de la thématique évoquée. Une des contraintes du projet est également de récupérer des données depuis une capture Web, c'est-à-dire en récoltant les données de sites Internet. Un de nos objectifs sera donc de trouver un site Internet sur lequel nous pourrons effectuer la capture.

Il sera également nécessaire de réaliser, en dehors des différents dashboards et métriques créées avec Dataiku, réaliser une modélisation du schéma en constellation de notre base de données multidimensionnelle, une modélisation conceptuelle et logique. Il faudra aussi définir les pré-traitements réalisés avec Dataiku.

**(PS - Toutes les illustrations évoquées dans le rapport n'y seront pas forcément présentes dans le rapport en raison de la taille maximale imposée. Merci de vous rapporter à la présentation ou au dossier contenant les images afin d'avoir accès aux illustrations. Les sources sont citées à la fin du rapport pour ne pas encombrer les pages).**

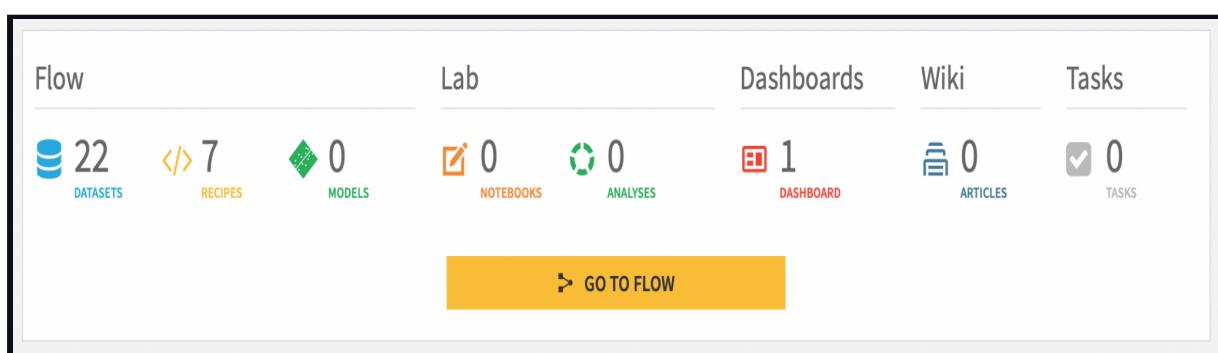
# Présentation de Dataiku et des données

Dataiku nous permet de construire notre projet en reliant les bases de données entre elles afin de mettre en avant de potentiels paternes. Une fois les données récoltées ou scrapées d'Internet, nous les ajoutons au projet et réalisons toutes les modifications que l'on juge nécessaires. à l'aide de "recipes", ou recettes, dans lesquelles on supprime les exemples inutiles, calcule de nouvelles colonnes. On peut également lier les tables à d'autres tables lorsqu'elles partagent des colonnes aux valeurs similaires (exemple, le code des pays), et ainsi augmenter en complexité et qualité le niveau de nos données. Nous avons décidé de nous concentrer uniquement sur l'année 2020 pour notre étude, puisque pour réaliser des comparaisons sur plusieurs années, nous aurions largement dépassé le temps de présentation impari.

Nous nous sommes donc principalement intéressés dans un premier temps aux facteurs économiques et de développement humain à travers le monde. Nous avons donc récupéré des données comme la répartition de la population mondiale, des salaires annuels moyens ou les taux de chômage pour les facteurs économiques. Pour représenter les indices de développement humain nous avons, au lieu d'utiliser l'IDH, utilisé des métriques comme l'accès à l'éducation primaire, à l'électricité ou au nombre de lits d'hôpitaux par millier d'habitants.

Pour illustrer la migration, nous avons récupéré des données comme par exemple le nombre de migrants internationaux par pays, le nombre d'émigrants par pays ou même le nombre de déplacements liés aux conflits et catastrophes locales. Les déplacements sont parfois engendrés non par le besoin de trouver un meilleur environnement mais par la nécessité de survivre.

Afin d'illustrer notre étude à l'échelle locale, nous avons pris l'exemple de la France. Nous nous sommes renseignés sur les origines des migrants en France et la proportion d'étrangers dans la population des pays de l'UE (Union Européenne) afin de réaliser la comparaison à l'échelle continentale. Afin de rentrer dans les détails à l'échelle locale, nous nous sommes intéressés à la situation des départements français.



Résumé de notre projet sur Dataiku (nombre de datasets, de recipes, ...)

D'un point de vue plus technique, nous avons enregistré les bases de données sur Dataiku. Une fois toutes les bases présentes sur l'application, nous avons commencé à réaliser des opérations de nettoyage. Sur toutes nos bases, nous avons réalisé la suppression de tous les 3-codes invalides (**Delete rows with invalid values**) et on transforme les données si nécessaire.

Souvent, il existe des lignes représentant non pas des pays mais des régions (exemple, Europe ou Maghreb). Ces informations ne nous intéressent pas puisqu'on ne peut pas effectuer de liaison de données (**Enrich -> Join** ou **Fuzzy-join** sur les 3-codes) pour en extraire de nouvelles.

**Fuzzy join** a été utilisé surtout pour les départements français, puisque l'orthographe change entre base de données, surtout au niveau de la ponctuation.

Par la suite, on ajoute les géo-localisations des pays grâce à la longitude et latitude des pays et à la transformation (**Geography -> Create GeoPoint from lat/lon**).

On utilise par la suite des formules mathématiques (**Math / Number -> Formula**) pour générer de nouvelles valeurs comme par exemple le PIB par habitant, depuis le PIB et le nombre d'habitants du pays.

## Schémas

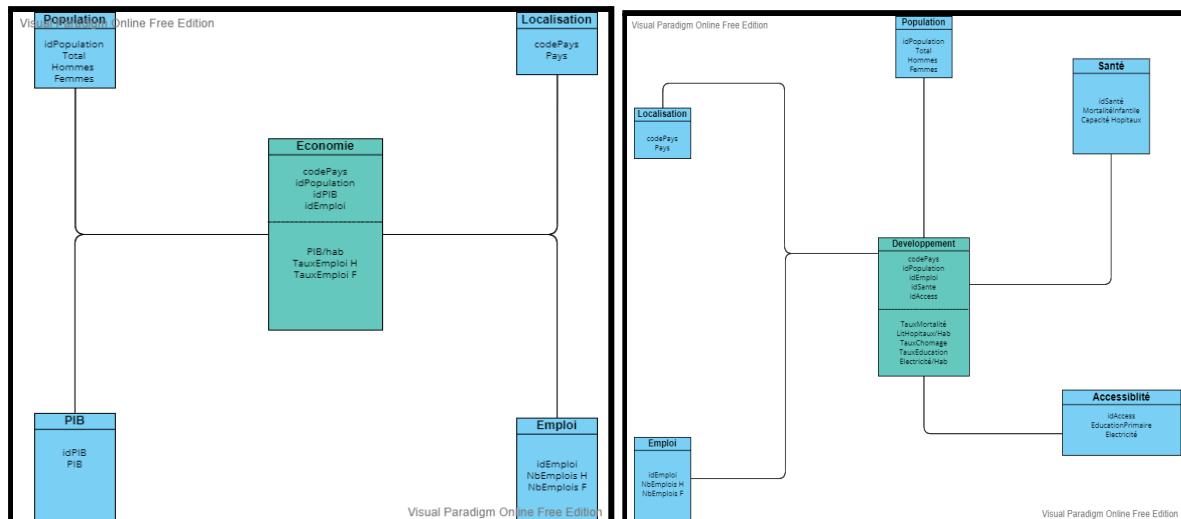


Schéma en étoile du fait Économie

Schéma en étoile du fait Développement

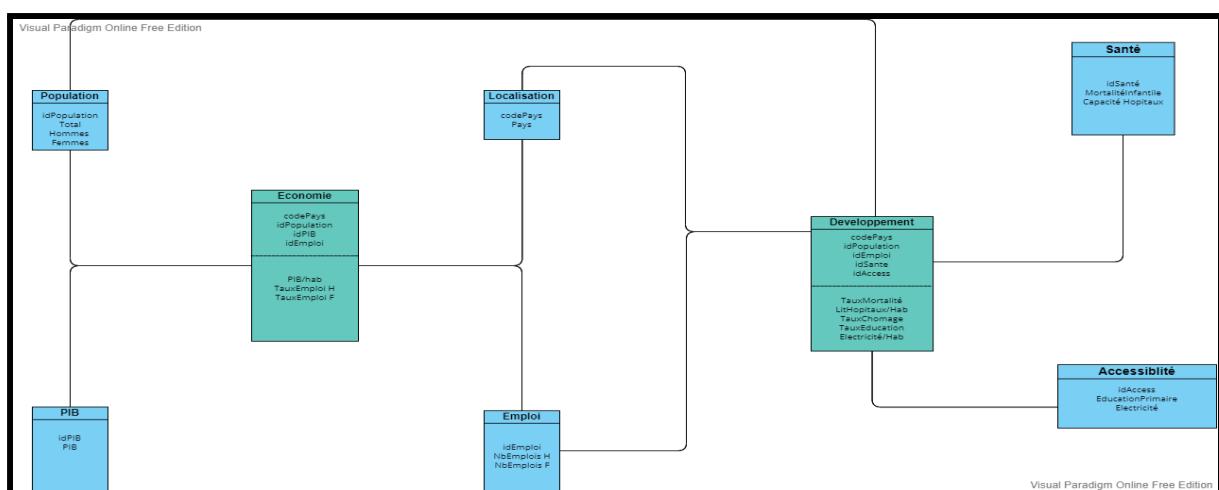


Schéma en constellation des faits

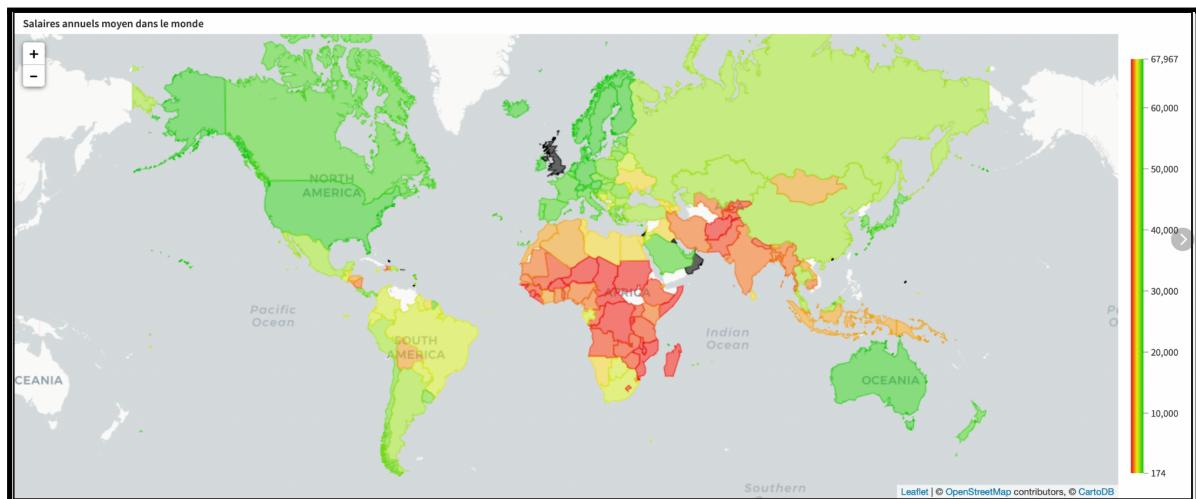
# Les facteurs économiques

Comme mentionné dans la partie précédente, on dispose de plusieurs métriques afin d'évaluer l'aspect économique des pays.

Afin de pouvoir représenter les valeurs sur une carte du monde, ce qui est plus pratique pour visualiser certaines données, nous avons dû rechercher une base de données contenant toutes les coordonnées géographiques de chaque pays afin de pouvoir former les localisations géographiques sur la carte. En utilisant les “recipes”, nous avons alors relié les localisations aux autres bases de données en utilisant le plus souvent le **3-code** des pays (FRA = 3-code de France).

On s'intéresse avant tout autre chose à la répartition de la population à travers le monde. Un pays ayant une forte population est plus susceptible d'apparaître dans des métriques concernant la migration. Le foyer de population le plus important étant l'Asie du Sud-Est, on remarque qu'il existe tout de même d'autres foyers comme l'Amérique du Nord. On estime alors retrouver les pays de ces régions dans les prochaines métriques lorsque l'on étudiera les déplacements plus en détail.

On étudie par la suite la répartition des salaires annuels moyens, en dollars, à travers le monde. On remarque que les salaires moyens sont faibles en Afrique, ainsi qu'en Asie du Sud-Ouest (Pakistan, Afghanistan, ...) tandis qu'on approche les 45.000 dollars annuellement en moyenne en Europe.



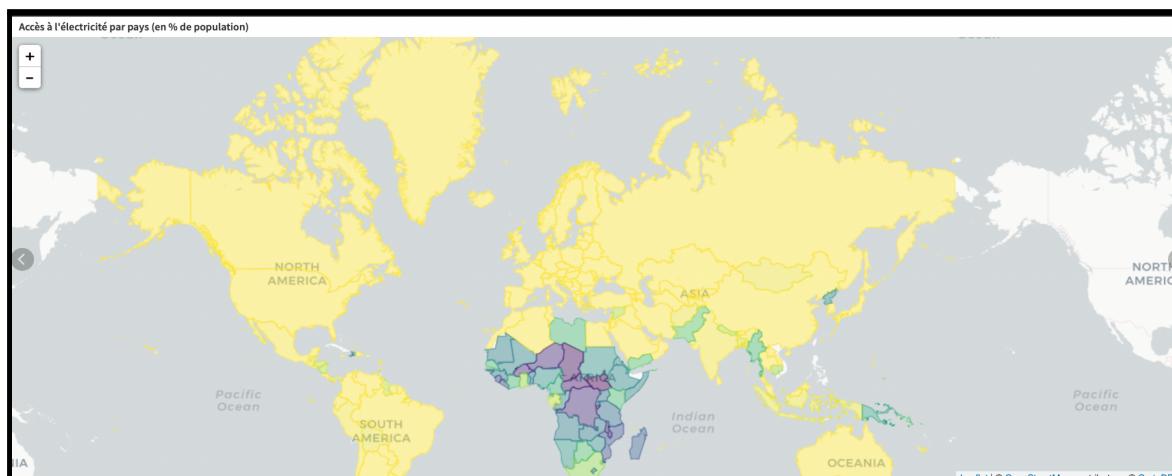
On dispose également des taux et des populations de chômage pour chaque pays. On remarque que la population la plus au chômage dans le monde est celle d'Afrique du Sud, avec à peu près 29% de la population sans emploi. Par contre, il ne s'agit pas de la population au chômage la plus importante au monde, puisqu'il s'agit de l'Inde où on compte plus de 110 millions de personnes au chômage. On fera donc attention à l'Inde lors de l'étude des migrations.

Nous avons également récolté le PIB des pays à travers le monde afin de pouvoir afficher la répartition globale et identifier les pays ayant la valeur la plus importante. Sans trop de surprises, il s'agit des Etats-Unis et de la Chine. Cependant, le PIB par habitant le plus important est celui de la principauté de Monaco.

# Les facteurs humains

Mais pourquoi est-ce que les populations se déplacent-elles? S'agirait-il plus d'un problème de qualité de vie locale plutôt que de rémunération salariale ou de surpopulation de l'environnement? On étudie alors quelques indices permettant la qualité de vie à travers le monde. Comme mentionné dans l'introduction, nous n'avons pas utilisé l'IDH, qui est une valeur assez vague, mais plutôt l'accès aux besoins primaires et au confort de vie de l'Homme dans les différents pays.

En premier lieu, nous nous intéressons à l'accès à l'électricité à travers le monde. Il s'agit aujourd'hui d'un besoin primaire puisque l'électricité nous permet de nous chauffer, nous nourrir et même de travailler grâce aux appareils électroniques. On remarque alors qu'en dehors de quelques pays en Asie ou Amérique Centrale, les populations ayant le moins accès à l'électricité sont celles d'Afrique Subsaharienne. On y retrouve là-bas les plus faibles taux d'accès du fait que la majorité des habitants de ces habitants se situent en zone rurale, où les installations nécessaires ne sont dans la majorité des cas absentes.



Nous étudions également quelques métriques liés à la jeunesse, plus particulièrement à l'accès à l'éducation pour les mineurs ainsi que les taux de mortalité infantile à travers le monde. Les données n'étant pas complètes puisqu'elles ne sont pas présentes pour un grand nombre de pays, on remarque tout de même une corrélation entre la mortalité infantile et l'accès à l'éducation pour les plus jeunes. La région du monde la plus affectée est une nouvelle fois l'Afrique Subsaharienne. Les infrastructures nécessaires afin de permettre le bon développement des enfants ne sont donc pas forcément présentes, comme les hôpitaux ou les écoles.

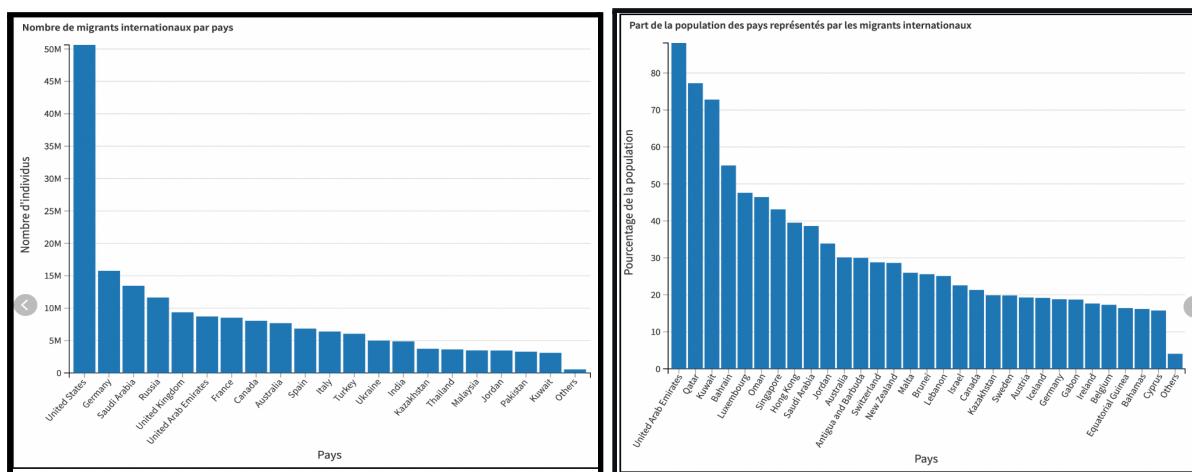
A propos des hôpitaux, nous avons également étudié le nombre de lits d'hôpitaux par 1000 habitants dans le monde. Cette fois-ci, nous ne pouvons pas étudier le cas de l'Afrique puisque pour la majorité des pays africains, les données sont manquantes. On remarque tout de même que les plus hautes valeurs se situent en Europe occidentale et en Asie de l'Est, le Japon et la Corée du Sud plus précisément.

# Les migrations en chiffres

Nous voilà maintenant au vif du sujet. Afin de pouvoir juger les déplacements de population à travers le globe, nous sommes dans un premier temps intéressés aux nombres de migrants par pays et la part de la population qu'ils représentent dans ces pays.

On remarque alors que le pays avec le plus grand nombre de migrants dans le monde est celui des Etats-Unis, avec près de 50 millions de migrants internationaux, suivi de l'Allemagne avec 15 millions. On retrouve la France avec à peu près 8 millions de migrants. Le rêve américain est donc toujours bel et bien réel puisqu'un nombre de personnes tentent leur chance chaque année, venant notamment d'Amérique Centrale et du Sud, dans l'espoir d'améliorer leurs conditions de vie.

Cependant, les pays dont les migrants représentent la plus grande partie de la population locale sont le Qatar et les Emirats Arabes Unis, avec environ 80% de la population locale représentée par les migrants. Ayant connu une croissance fulgurante ces dernières décennies grâce aux pétroles, ces pays sont aujourd'hui plein d'opportunités de travail.



Nombre de migrants par pays

Pourcentage de la population représenté par les migrants

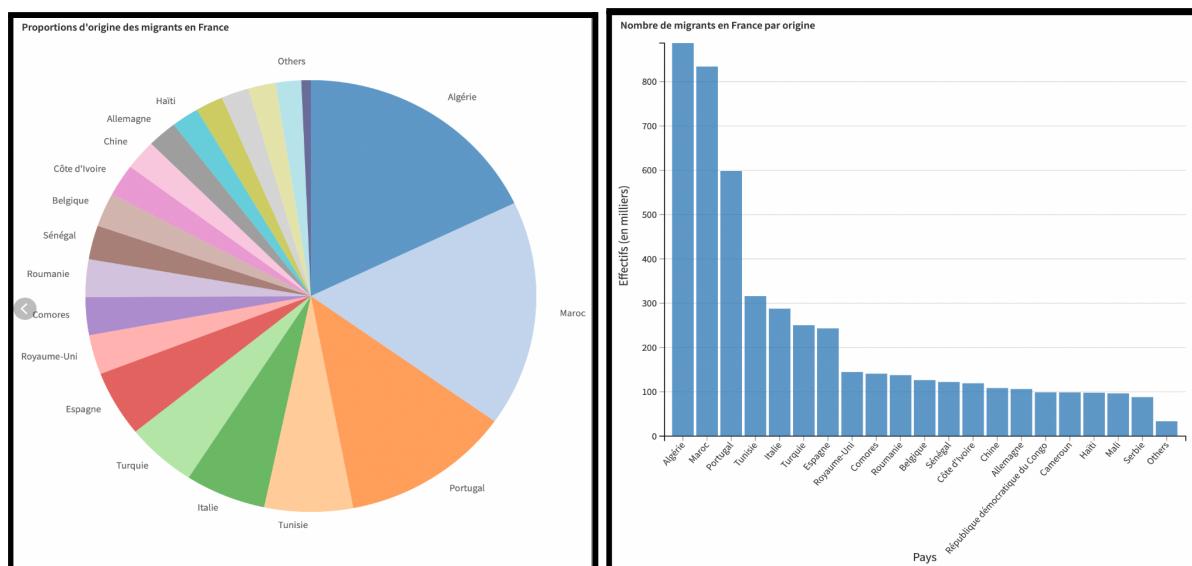
Par la suite, nous avons récolté des données à propos d'émigrants par pays et le nombre de migrants liés aux conflits et catastrophes locales. Sans surprise, on retrouve l'Inde en première position du nombre d'émigrants avec plus de 17 millions de personnes ayant quitté le pays. De ce fait, tous les facteurs évoqués précédemment ont permis d'expliquer ce nombre. On retrouve en Afghanistan le plus grand nombre de déplacements liés aux conflits et catastrophes locales. La cause principale de ces déplacements est le groupe des talibans, impliqué dans plusieurs conflits locaux, laissant de nombreuses familles sans protection et parfois même sans ressources.

En termes d'arrivées et de sorties des pays, on retrouve les Etats-Unis, l'Allemagne et la Turquie en tête du classement des arrivées sur le territoire tandis qu'au niveau des sorties, on retrouve étonnamment le Vénézuela devant l'Inde. Le Vénézuela est en crise depuis quelques années, à cause de la corruption politique, du chômage et du manque de ressources et s'en est donc suivi un exode massif des vénézuéliens.

# Etude locale - Le territoire français

Dans le cadre de notre projet, nous sommes amenés à étudier notre problématique à l'échelle locale. Nous avons alors décidé d'étudier le cas de la France, puisque premièrement, nous vivons en France, et que deuxièmement, le sujet des migrants fait souvent polémique en France depuis maintenant quelques décennies. Nous nous sommes alors intéressés plus particulièrement à la répartition de l'origine des migrants en France, de la répartition des populations au sein des départements français ainsi qu'aux nombres de migrants dans le reste des pays de l'UE.

De ce fait, on découvre que les migrants en France viennent principalement du Maghreb, plus précisément du Maroc et de l'Algérie. Il existe une relation importante entre la France et ces pays puisque le Maroc a été sous protectorat français entre 1912 et 1962 et l'Algérie a été une colonie française entre 1830 et 1962. Il s'agit donc des traces du passé, puisque de nombreuses familles maghrébines ont décidé de s'installer en France et continuent jusqu'à aujourd'hui de le faire.



**Répartition de l'origine des migrants en France**

En Europe, les pays où les migrants représentent une part importante de la population sont le Luxembourg, Malte et l'Allemagne. Le Luxembourg, du haut de ses 600 000 habitants, est un pays intéressant où travailler puisque comparé à la France, le salaire minimum est de 2100 euros par mois, contre 1500 euros en France.

**Nombre de migrants en fonction de l'origine en France**

Rentrions maintenant dans les détails et discutons des départements français. Les départements où les migrants représentent à la fois la plus grande partie de la population mais également les plus grandes populations de migrants dans les départements français sont la Seine-Saint-Denis, le Val-de-Marne et Paris. La Guyane comptabilise également un grand nombre de migrants puisqu'ils représentent 30% de la population. Dans le cas de la Guyane, les migrants y sont d'origine du Brésil ou de Haïti notamment.

# Conclusion et idées

Nous avons donc mis en avant plusieurs facteurs permettant d'expliquer les mouvements à l'échelle internationale, avec notamment des facteurs économiques ou de développement humain.

Il est aujourd'hui de plus en plus dur de trouver un emploi stable notamment dû à la compétition pour les postes, qui est encore plus féroce dans les pays à la population importante. D'autres facteurs comme l'accès à l'éducation ou aux soins permet également de justifier ces déplacements mais il faut réaliser une chose. Les populations des pays où la situation humaine est déplorable aimeraient vivre ailleurs mais par faute de moyens, elles ne peuvent quitter leur pays.

Un triste exemple est celui des migrants africains effectuant un long voyage, malheureusement beaucoup trop souvent mortel, depuis les pays d'Afrique Subsaharienne jusqu'à la Libye puis naviguent sur des embarcations de fortune en direction des côtes européennes.

Il faut donc mettre en place les dispositifs nécessaires pour accueillir les populations en provenance de l'étranger. On s'intéresse désormais uniquement au cas de la France pour la proposition d'idées d'implémentation de business ou de structures.

Une première idée serait de, en fonction des populations de migrants dans les différents départements (on ne possède aucun compte exact, il faudrait effectuer une étude supplémentaire pour obtenir les valeurs), ouvrir des boutiques et magasins vendant des biens ou de la nourriture en provenance des pays d'origine des migrants. On cherche souvent à se reconnecter à nos origines, quel que soit le moyen, et cela représente donc une possible idée de business.

La seconde idée serait de mettre en place des dispositifs, des structures, facilitant les personnes en provenance de l'étranger les plus défavorisées à trouver un travail et faire en sorte que leurs enfants, s'ils en ont un accès à l'éducation publique. Il existe aujourd'hui déjà plusieurs dispositifs de ce genre mais la population mondiale est en croissance et la France étant un pays développé attirant un grand nombre de migrants, il est donc primordial de s'assurer des conditions d'accueil des migrants.

## Sources et liens vers les bases de données et articles

**Guyane et migration :**

<https://www.histoire-immigration.fr/dynamiques-migratoires/la-guyane-ou-la-chronique-d-une-mondialisation-migratoire>

**Taux Chômage - Chômage par Pays :**

<https://databank.worldbank.org/source/world-development-indicators>

**Education concernant les enfants :**

<https://databank.worldbank.org/source/world-development-indicators>

**Latitudes et Longitudes des Pays :**

<https://www.kaggle.com/datasets/paultimothymooney/latitude-and-longitude-for-every-country-and-state>

**Mortalité infantile :**

<https://databank.worldbank.org/source/world-development-indicators>

**Immigration en France entrants et sortant :**

<https://www.insee.fr/fr/statistiques/3633212#onglet-1>  
<https://www.ined.fr/fr/tout-savoir-population/chiffres-france/immigres-etrangers/>

**Population des Pays :**

<https://databank.worldbank.org/source/world-development-indicators>

**Population en France :**

<https://www.insee.fr/fr/statistiques/1893198>

**Salaire moyen par Pays :**

[https://www.ilo.org/shinyapps/bulkexplorer52/?lang=fr&segment=indicator&id=EAR\\_4MTH\\_SEX\\_OC\\_U\\_CUR\\_NB\\_A](https://www.ilo.org/shinyapps/bulkexplorer52/?lang=fr&segment=indicator&id=EAR_4MTH_SEX_OC_U_CUR_NB_A)

**Salaire minimum par Pays :**

[https://www.ilo.org/shinyapps/bulkexplorer52/?lang=fr&segment=indicator&id=EAR\\_4MTH\\_SEX\\_OC\\_U\\_CUR\\_NB\\_A](https://www.ilo.org/shinyapps/bulkexplorer52/?lang=fr&segment=indicator&id=EAR_4MTH_SEX_OC_U_CUR_NB_A)