

Projet Données sur le Web

Open Data

Le milieu carcéral

SOMMAIRE

Introduction	3
Gestion du projet	4
Composition du groupe	4
Organisation temporelle	5
Présentation des données	6
L'Open Data	6
Présentation	6
L'échelle de Tim Berners-Lee	6
Explication des données	7
Population carcérale des pays	7
Nombre d'infractions enregistrées	7
Taux de chômage	8
PIB par habitant	8
Coefficient de Gini	9
Diagramme de classes	10
Analyse des données	11
Population carcérale pour 100 000 habitants	12
Infractions enregistrées pour 100 000 habitants	13
PIB par habitant	14
Taux de chômage	15
Coefficient de Gini	16
Présentation des résultats	17
Allemagne	17
Autriche	17
Espagne	17
La France	18
Grèce	18
Portugal	18
Argumentation finale	19
Applications existantes proches	20
Problèmes sur les données	20
Une application	21
Présentation des traitements	25
Conclusion	29
ANNEXES	A
Glossaire	A
Webographie	D

1. Introduction

Pour commencer, ce projet nous a donné l'occasion d'exploiter des données dites Open Data sur un thème libre. L'objectif était de croiser plusieurs jeux de données et de montrer la corrélation (ou non) entre ceux-ci.

Dans les parties qui vont suivre nous aborderons :

- La gestion de notre projet qui peut se décomposer de façon plus claire : Nous verrons tout d'abord la composition de notre groupe. Ensuite, nous transitionnerons vers notre démarche de travail. Enfin nous terminerons cette partie en établissant notre organisation temporelle.
- La présentation de notre travail où nous reviendrons sur le concept de l'Open Data en premier lieu. La suite de cette partie sera constituée d'une argumentation la plus précise quant au choix de nos données.
- Une analyse de nos données. C'est notamment dans cette partie que nous montrerons la corrélation entre les différentes données que nous comparons. Nous pensons que cette partie représente la partie la plus importante de notre projet, c'est pourquoi nous souhaitons parler des différents problèmes encourus lors de la manipulation des données.
- Une présentation de notre application ainsi que de ses fonctionnalités. Cette partie servira principalement à donner une idée générale de son développement (langage utilisé, nous placerons également quelques captures d'écran pour en montrer l'interface graphique ainsi que les courbes et les données présentes dessus).
- Enfin nous reviendrons sur notre expérience de façon succincte dans une conclusion.
- En annexe, nous proposerons un glossaire ainsi qu'une webographie pour apporter des informations pertinentes à la lecture de notre rapport. Tous les mots ou les groupes de mots écrits en *italique* suivi du symbole “*” apparaîtront dans ce glossaire (classé par ordre alphabétique).

Cette introduction permet de donner un plan de lecture de notre rapport. Dans un premier temps, nous allons développer la partie “Gestion du projet”.

2. Gestion du projet

2.1. Composition du groupe

Pour commencer, nous sommes un binôme d'étudiants : Luca GARIC et Michaël LOPES. Afin d'obtenir une organisation optimale dans notre groupe, nous avons décidé de nous répartir les tâches de façon équitable. En effet, nous avons travaillé ensemble pour choisir le sujet à développer. De plus, nous avons défini ensemble la portée des données (les pays analysés dans le sujet) et le type des données que nous souhaitons utiliser dans notre sujet (comme la nature des indicateurs que nous voulions exploiter).

Après avoir délimité notre sujet, nous avons décidé de nous répartir les tâches pour être le plus efficient possible. En effet, Michaël s'est occupé de récolter la base de notre projet qui correspond aux différentes populations carcérales des pays concernés. De plus, ses compétences de montage vidéo ont été utilisées pour la réalisation de la nôtre, il a donc travaillé en majorité dessus. Luca s'est quant à lui retrouvé à chercher les indicateurs potentiellement pertinents au sujet comme le PIB par habitant, le taux de chômage ou le coefficient de Gini avec les *infractions** commises (avec la coopération de Michaël sur cette dernière), ses capacités rédactionnelles ont été mises en œuvre sur la création du rapport.

Ensuite, nous avons croisé nos données récoltées ensemble afin d'obtenir deux avis distincts sur les résultats obtenus, cela nous a notamment permis d'obtenir des certitudes ou non quant à la pertinence de certains jeux de données. Enfin, l'application a été réalisée avec la coopération des deux membres. Bien sûr, nous tenons à préciser que nous n'étions pas cantonnés à nos tâches respectives, nous nous sommes déjà entraides malgré la définition de nos travaux.

2.2. Démarche de travail

Tout d'abord, la justice est un thème important et très intéressant à étudier. Nous étions tous les deux d'accord pour mettre au point un sujet qui pouvait s'en approcher ou d'au moins élaborer un travail qui pouvait montrer des jeux de données en lien avec ce thème. Nous avons donc notre sujet de base. La prochaine étape fut de discuter des types de données que nous voulions utiliser. En premier lieu, nous avons trouvé un document PDF récapitulant l'ensemble des détenus en France ainsi que plusieurs paramètres intéressants, comme la répartition des sexes en prison ou encore les détenus en attente de jugement. Cependant, nous n'avions que ces données à disposition et aucun autre jeu de données à croiser. Malheureusement, nous avons donc dû rectifier légèrement notre ligne de route concernant notre sujet.

Après un temps conséquent de recherches supplémentaires, nous avons trouvé les populations carcérales d'une majorité des pays développés ou en voie de développement dans le monde, et même sans données plus précises pour ces pays, nous savions que nous avions trouvé notre nouvelle base de travail. Dans le même temps, nous avons fait une découverte intéressante et sans doute motrice pour le reste de notre projet : nous avons trouvé le nombre d'infractions commises par chaque pays d'Europe et même si certains

pays étaient incomplets, nous avons tout de même réussi à définir un périmètre de pays sur lequel travailler. Suite à cette trouvaille, nous nous sommes posés la question suivante “Qu’est ce qui peut potentiellement motiver ces infractions ?”. C’est à ce stade que nous avons défini notre sujet final : Comparer le nombre d’infractions / *population carcérale** en Allemagne, Autriche, Espagne, France, Grèce et Portugal avec des indicateurs tels que le PIB par habitant qui permet de mesurer le niveau de vie des habitants et le taux de chômage des pays qui permet d’avoir une vue sur la situation sociale de ceux-ci. Enfin, comparer le coefficient de Gini avec les deux données ci-dessus permet de soulever l’influence des inégalités de richesses sur la criminalité.

Ce choix des pays s’est fait sur deux critères :

- Nous souhaitions rester en Europe avec des pays d’une situation économique plus ou moins proche de celle de la France
- Nous souhaitions avoir des pays ayant les données les plus complètes possible.

Concernant nos outils de travail, nous avons opté pour Discord en discussion instantanée et vocale, car il s’agit d’un logiciel que nous maîtrisons et qui s’avère pratique dans le contexte actuel. Ensuite, pour partager et stocker notre travail, nous avons utilisé Google Drive. Enfin, Excel a principalement servi à générer nos données récoltées ainsi que les graphiques associés.

2.3. Organisation temporelle

Tout d’abord, nous avons en tête d’obtenir un sujet pertinent pour le 30 septembre environ. Ce fut chose faite car la validation de notre sujet s’est faite le 28 septembre. Notre travail concret de recherche a réellement été initié à partir du 24 octobre environ, cela s’explique par d’autres priorités qui nous ont réquisitionnés avant cette date. À partir du 24 octobre et pendant deux semaines, nous avons été fortement actifs sur le projet. C’est notamment durant cette période que nous avons effectué l’essence de notre travail (c’est-à-dire la récolte des données ainsi que la mise en place des graphiques correspondant et le croisement de ceux-ci).

Nous sommes revenus assez souvent sur notre rapport pour mettre à jour nos idées et obtenir un fil conducteur sur le travail restant. La rédaction du rapport a donc nécessité un travail assidu tout au long du projet.

Ensuite, nous avons réalisé un diaporama servant de support à notre présentation. L’objectif était d’obtenir des “slides” claires et concises reprenant la quintessence de notre sujet. Nous pensons terminer le diaporama pour la semaine du 16 novembre.

L’application fut commencée dès la fin de la récolte des données et de leur croisement (une fois le sujet complet), comme date de fin nous avons respecté la deadline officielle.

La vidéo fut réalisée après le diaporama explicatif de notre sujet, nous nous sommes fixé d’obtenir une vidéo correcte à environ une semaine de la deadline officielle pour pouvoir prendre du recul sur celle-ci et la perfectionner en cas de besoin.

3. Présentation des données

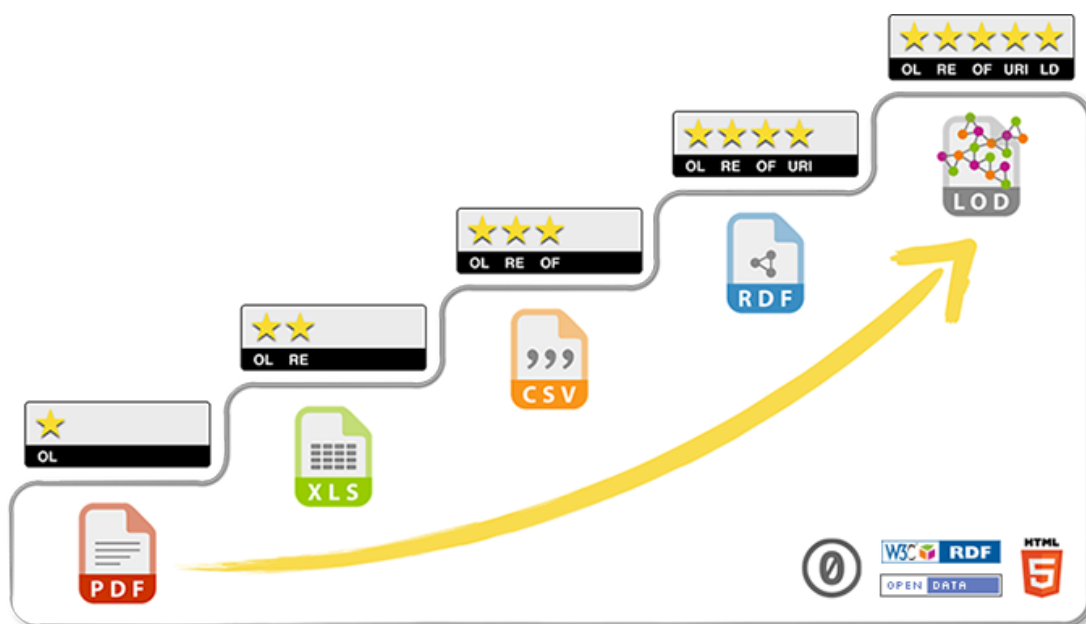
3.1. L'Open Data

3.1.1. Présentation

Tout d'abord, l'Open Data (donnée ouverte) représente des données pouvant être partagées et accessibles par tous. Tout le monde peut aussi y contribuer et les réutiliser. Ces données sont brutes, complètes, libres de tout *copyright** et non discriminantes.

3.1.2. L'échelle de Tim Berners-Lee

D'après l'inventeur du World Wide Web, les données se jugent de cette façon :



- ★ Publiée sur le Web (peu importe leur format) avec une licence ouverte
- ★★ Publiée en tant que données structurées (ex : document Excel et non une image)
- ★★★ Publiée dans un format ouvert et non-propriétaire (ex : CSV* et non un Excel)
- ★★★★ Utilisation des *URI** pour désigner des choses dans les données, afin que les utilisateurs puissent faire des références à celles-ci
- ★★★★★ Liée à d'autres données pour y ajouter du contexte

Lors de notre expérience et durant nos recherches, nous avons trouvé des données structurées dans des formats non-propriétaires tels que le format CSV, nous avons aussi trouvé des données répertoriées sous format Excel. Les données que nous exploitons se rangent en moyenne entre la deuxième et la troisième catégorie. Nous avons aussi un jeu de données en format PDF, on peut donc le classer en catégorie une.

3.2. Explication des données

Dans un premier temps, nous avons recherché un maximum de données pertinentes et liées à notre jeu de données : la population carcérale. Toutes les sources des jeux de données sont disponibles dans notre partie webographie (annexe page D). Nous avons répertorié certains mots-clés dans une partie glossaire disponible à la page A de l'annexe également.

3.2.1. Population carcérale des pays

Nous nous sommes concentrés sur des données pouvant avoir une influence assez vive sur les populations carcérales. Tout d'abord, les populations carcérales sont calculées pour 100 000 habitants, cela s'explique par une différence du nombre des populations dans les différents pays analysés. Cela permet aussi de donner un sens réel à notre analyse tout en restant précis.

Les données récoltées proviennent du site ec.europa.eu qui peut s'identifier comme étant le site des statistiques de l'Union Européenne. Cette entité est la source (en collaboration avec le Système de Statistique Européen) et le diffuseur de ces données. Le thème de la "population et des conditions sociales" peut être affecté à ce jeu de données. Les valeurs les plus anciennes datent de 2008 tandis que les plus récentes sont de 2018. Le jeu de données entier contient 1508 valeurs (la population carcérale de 41 pays (nombre de colonnes) sur 11 années (nombre de lignes) sont recensées). Même si nous n'avons aucune indication sur la date de création de ce jeu de données, nous savons néanmoins que la fréquence de mise à jour est annuelle, et que la dernière mise à jour a eu lieu le 10 juillet 2020. Les données sont sous la "*licence Creative Commons** Attribution 4.0 International, © Union européenne, 1995-2020". Le format des données est assez portable dans le sens où plusieurs de ceux-ci sont disponibles en téléchargement : *XLS**, *SDMX**, *XML**, *JSON**, *CSV*, *HTML**, *PC AXIS**, *SPSS**, *TSV** et enfin *PDF**. La taille totale de ces données s'élève à 34.2 Ko. Ce site nous permet de sélectionner les données voulues avant de les télécharger. Donc, elles pèsent 15 Ko après sélection.

Malheureusement, des données sont vides concernant certains pays et certaines années, nous n'avons pas pu nous concentrer sur tous les pays que nous voulions à la base faute de données renseignées.

3.2.2. Nombre d'infractions enregistrées

Ensuite, il fut assez évident de s'orienter sur le nombre d'infractions des pays. En effet, les infractions sont sujettes à des peines de prison et peuvent directement influencer les populations carcérales. Comme pour celles-ci, les nombres d'infractions seront donnés sous le format d'infractions pour 100 000 habitants, les raisons sont les mêmes que celles expliquées plus haut.

Nous retrouvons le site Eurostat qui représente la source de ces données. Ici, les données portent sur le thème de la "criminalité et de la justice pénale". Concernant la dimension des données, l'ensemble recense 11 années (nombre de colonnes) et 41 pays (nombre de

lignes). Comme pour le jeu de données précédent, les données les plus anciennes recensées sont en 2008 pour finir en 2018. Le jeu de données total propose 10626 valeurs (un nombre assez conséquent qui se traduit par un large panel d'infractions différentes : *vols**, *vols qualifiés**, *homicides intentionnels**, *enlèvements**, *agressions**, *violences sexuelles**, *cambriolages** et *actes illicites liés au domaine de la drogue**). La date de création n'est pas renseignée. Nous savons néanmoins que la fréquence de mise à jour est annuelle et le dernier rafraîchissement des données s'est effectué le 20 juillet 2020. Les données sont sous la "licence Creative Commons Attribution 4.0 International, © Union européenne, 1995-2020". Le site propose un large choix de formats à télécharger (XLS, CSV, HTML, PC AXIS, SPSS, TSV et PDF). La taille totale des données fait 166 Ko. Ce site nous permet de sélectionner les données voulues avant de les télécharger. Donc, elles pèsent 27.5 Ko après sélection.

Comme pour les données précédentes, certaines années ne sont pas renseignées (notamment pour la France en 2018) ce qui a compliqué l'exploitation des données.

3.2.3. Taux de chômage

De plus, nous avons choisi de montrer le taux de chômage des pays qui indique le pourcentage des actifs qui sont au chômage. Cet indicateur peut montrer certaines tensions sociales dans les pays.

La source des données reste le site Open Data de l'Europe. Le thème abordé est le "social et travail". Les données les plus anciennes datent de 2008 et se terminent en 2019. La dernière mise à jour des données s'est effectuée le 07/10/2020. Le tableau comporte 12 colonnes (12 années) ainsi que 41 lignes (41 pays différents). Le nombre de données total sans case vide est de 1371. Nous n'avons malheureusement aucune information sur la date de création. Les données sont sous la "licence Creative Commons Attribution 4.0 International, © Union européenne, 1995-2020". Les formats disponibles au téléchargement sont XLS, SDMX, JSON, CSV, HTML, TSV et PDF. De plus, nous pouvons sélectionner nos données avant téléchargement, de ce fait la taille totale du fichier est de 28.8 Ko et la partie prélevée pèse 12.2 ko.

3.2.4. PIB par habitant

Ensuite, nous pensons que le niveau de vie des habitants peut influencer la criminalité et donc la population carcérale des pays. Pour cela, nous avons donc choisi d'analyser le PIB par habitant pour les pays concernés.

La source des données est "ec.europa.eu". Le thème ici représente l'économie et le niveau de vie des habitants. Les données les plus anciennes datent de 2000 et les plus récentes datent de 2019. Le document contient 20 colonnes et 38 lignes. La valeur du jeu de données total s'élève à 1492 entrées. Comme pour au-dessus, nous n'avons pas d'informations concernant la date de création du jeu et sa fréquence de mise à jour. Nous savons néanmoins que la dernière mise à jour a été faite le 13 novembre 2020. Ces données sont sous le format "licence Creative Commons Attribution 4.0 International, © Union

européenne, 1995-2020". Le site propose un large choix de formats à télécharger (XLS, SDMX, JSON, CSV, HTML, TSV et PDF). La taille totale des données fait 27,7 Ko. Ce site nous permet de sélectionner les données voulues avant de les télécharger. Donc, elles pèsent 12,1 Ko après sélection.

3.2.5. Coefficient de Gini

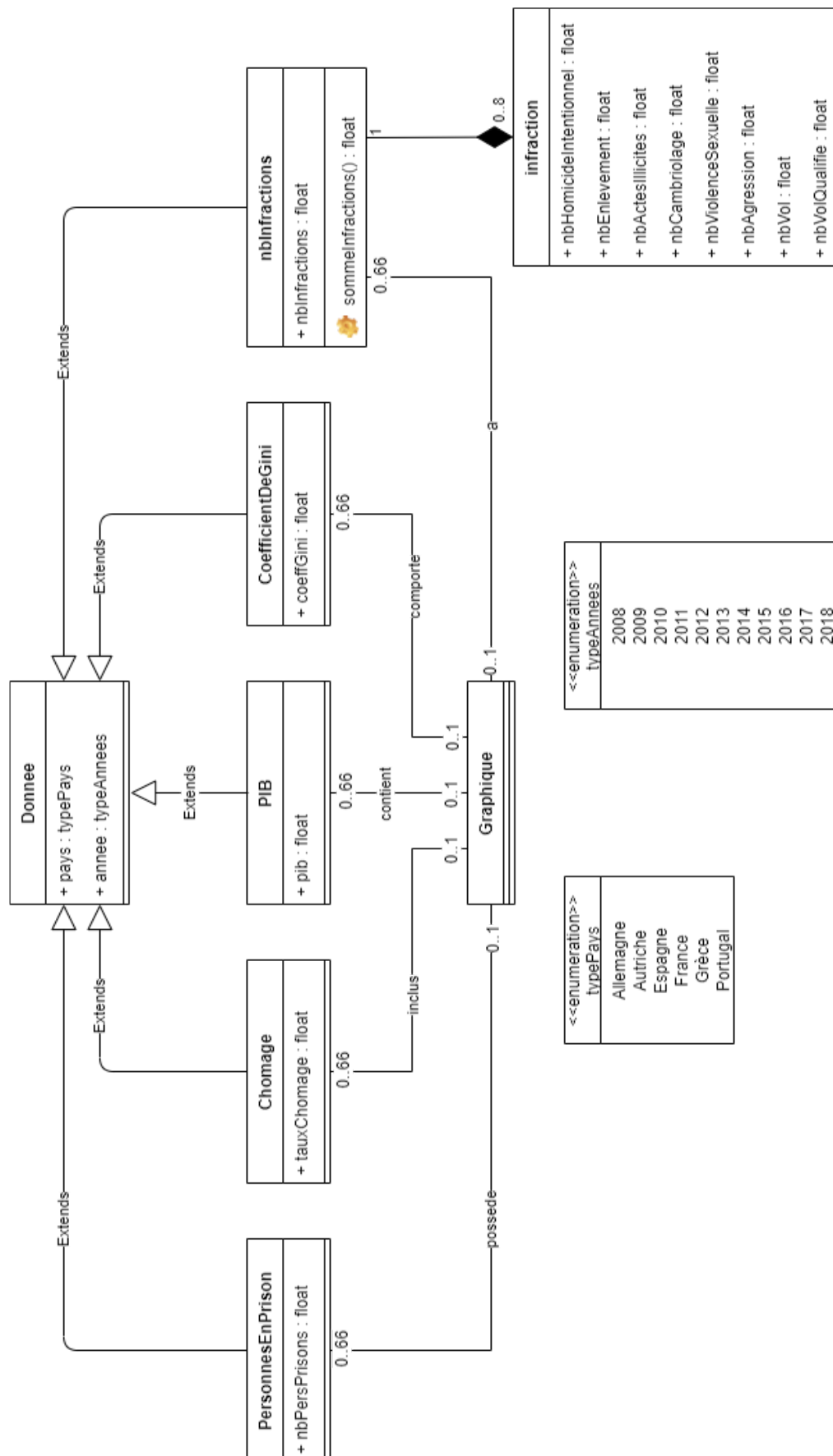
Enfin, le coefficient de Gini va permettre de nous montrer que les inégalités entre les différentes richesses des pays peuvent potentiellement impacter le nombre de prisonniers et le taux de criminalité.

La source des données provient toujours du même site "ec.europa.eu". Les données les plus anciennes recensées datent de 2008, les plus récentes datent de 2019. La dernière mise à jour a été effectuée le 27/10/2020. Le nombre total de valeur est de 448 entrées. Le document total comporte 41 lignes (car il y a 41 pays) et 12 colonnes (car les données sont indiquées sur 12 années). Nous ne pouvons pas renseigner de date de création. La licence dont les données sont soumises est : "licence Creative Commons Attribution 4.0 International, © Union européenne, 1995-2020". Les différents formats proposés lors du téléchargement sont les suivants : (XLS, SDMX, JSON, CSV, HTML, TSV et PDF). Enfin, nous avons l'opportunité de sélectionner les données à télécharger, donc nous pouvons dire que le poids total des données est de 46.5 Ko, et la taille du document après sélection fait 11.9 Ko.

D'une manière globale, les données que nous avons trouvées sont plutôt "chaudes". La seule exception que nous pouvons mentionner a été le fait de calculer le nombre d'infractions par pays, où certaines données se sont révélées être peu actualisées. Cela nous a poussé à utiliser une *courbe de tendance polynomiale** comme nous le montrerons plus bas. Concernant les autres jeux de données, la fréquence de mise à jour est active et les dernières actualisations sont récentes.

Aucune donnée n'est complexe, nous nous sommes juste basés sur différentes infractions pour calculer les infractions totales (vols, vols qualifiés, cambriolages, homicides, violences sexuelles, agressions, enlèvements et actes illicites faisant intervenir des drogues contrôlées ou des précurseurs).

3.3. Diagramme de classes



Nous avons choisi de concevoir notre modèle conceptuel à l'aide d'un diagramme de classes. Nous avons donc modélisé la structure de notre projet avec les différentes données qui composent les jeux choisis. Nous avons créé une classe infraction qui regroupe les types utilisés de ceux-ci. L'addition de ces infractions va résulter dans un total appelé "nbInfractions" d'où la présence d'une composition entre ces deux classes (si on supprime "infraction", "nbInfractions" n'existera plus). Pour résumer, nous avons aligné toutes les données utiles à nos croisements (le nombre de personnes en prison, le chômage, le PIB par habitant, le coefficient de Gini et donc le nombre d'infractions). Une donnée peut être l'un de ces champs d'où la présence des "extends" des indicateurs vers la classe "Donnée". Cette classe contient deux énumérations qui sont les années et les pays (les pays et les années sont fixes pour toutes les données). Enfin, un graphique contient 0 à 66 données pour chaque indicateur.

4. Analyse des données

La criminalité est un sujet grandissant non seulement en Europe mais aussi dans tous les pays du monde. Crises sociales et économiques s'enchaînent dans les pays, mais qu'en est-il de leurs impacts sur leur criminalité ? C'est précisément pour répondre à cette problématique que nous avons choisi ce sujet et nos indicateurs.

Concrètement, l'intérêt principal de notre projet est de démontrer les influences de certains mécanismes de la société sur la criminalité dans des pays d'Europe occidentale. Un de nos objectifs est de démontrer que chaque pays est affecté différemment en fonction de sa conjoncture économique et sociale. En effet, un pays peut être plus touché par les tensions sociales en son sein alors qu'un autre pays verra sa criminalité grandement influencée par le niveau de vie de ses habitants. Forcément, nous essayons de démontrer nos objectifs dans la portée que nous offre l'Open Data, certains indicateurs peuvent influencer la criminalité mais nous ne les affichons pas car ils ne sont pas disponibles en Open Data. Nous en parlerons plus en détails durant notre développement.

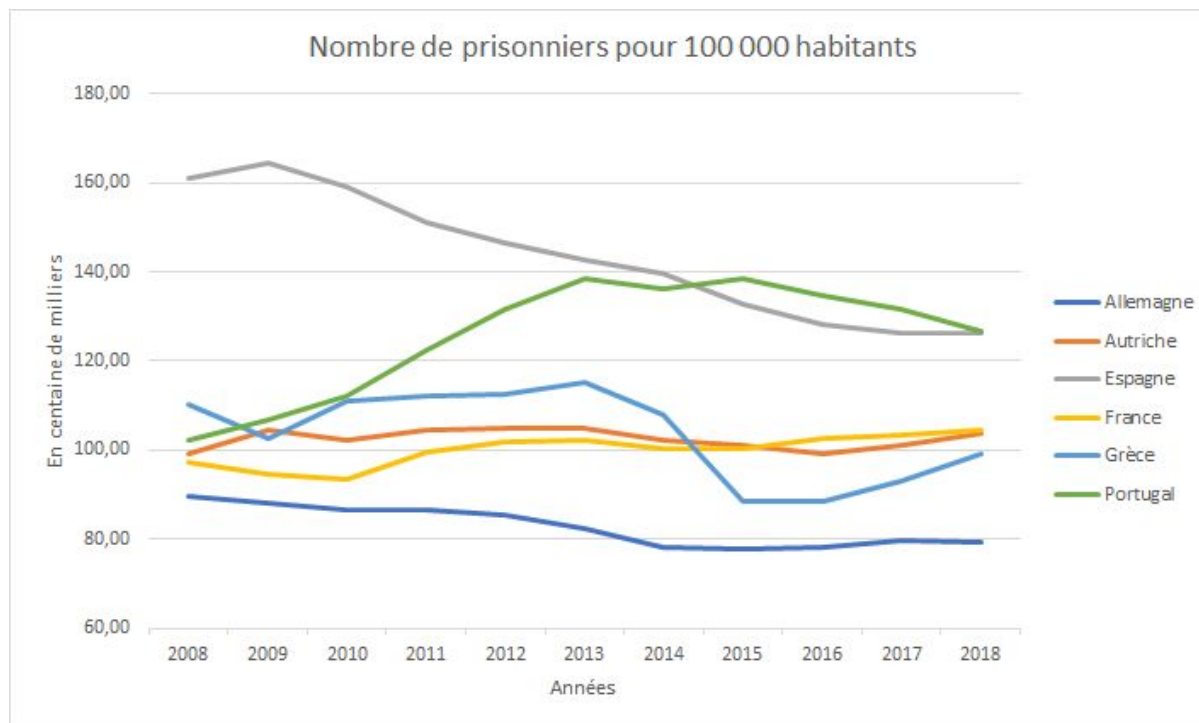
Nous avons choisi le taux de chômage car, pour nous, il représente un indicateur social pertinent car il expose la potentielle inactivité au sein d'une population et donc peut révéler certaines tensions parmi cette dernière. Nous essayons donc de faire corrélérer ce taux avec les infractions commises dans les pays d'Europe.

Pour commencer, nous allons donc vous présenter les populations carcérales en : Allemagne, Autriche, Espagne, France, Grèce et Portugal ainsi que leur nombre d'infractions.

Ces données seront croisées avec leur PIB par habitant, ainsi qu'avec leur taux de chômage et leur coefficient de Gini.

4.1. Population carcérale pour 100 000 habitants

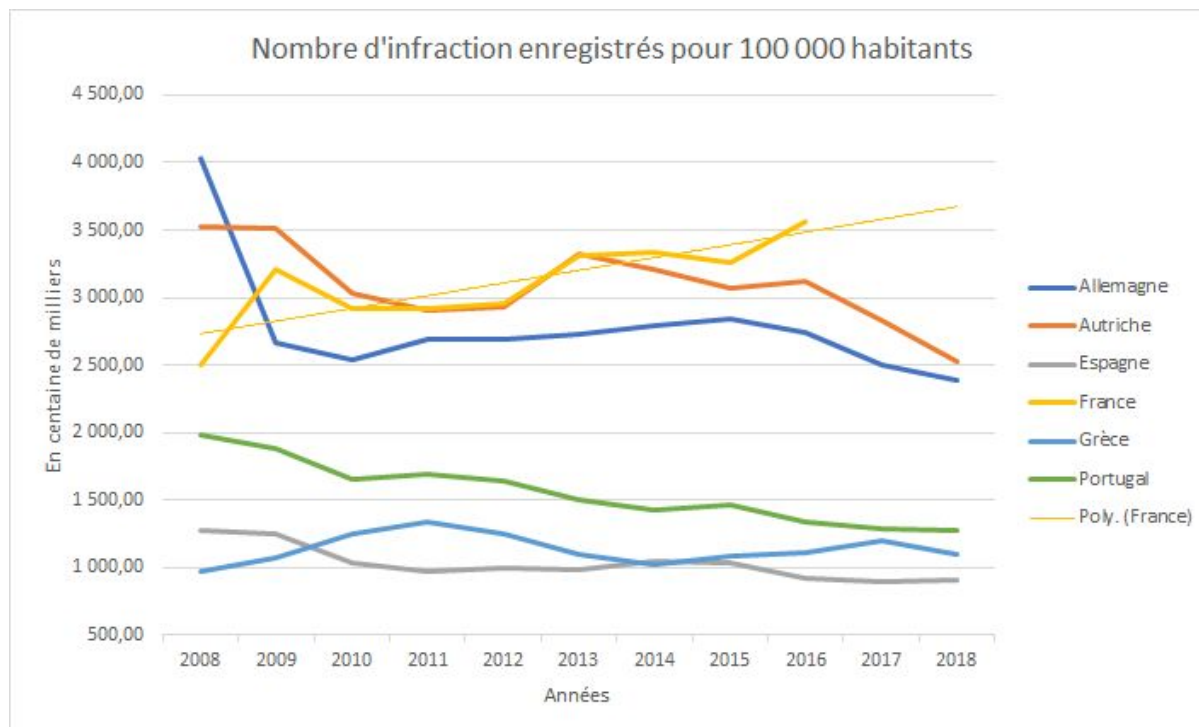
Dans un premier temps, nous souhaitons initialiser nos jeux de données avec les populations carcérales des pays choisis. Le graphique ci-dessous sert surtout à vous donner une idée des pays ayant une population carcérale élevée ou faible.



Nous pouvons déjà remarquer que l'Allemagne possède le nombre de prisonniers le plus faible (pour 100 000 habitants). La courbe diminue légèrement, elle commence donc à environ 90 prisonniers pour 100 000 habitants en 2008 pour arriver à environ 80 prisonniers dix ans après. C'est en Espagne et au Portugal que l'on peut retrouver le nombre de prisonniers le plus élevé où on retrouve environ 127 prisonniers pour 100 000 habitants. Cependant, nous pouvons observer une amélioration en Espagne qui partait avec 161 prisonniers en 2008 tandis que le Portugal en possédait 100 pour la même date. L'Autriche, la France et la Grèce ont des données assez similaires en 2018 : respectivement 104, 105 et 100, et à l'exception de la Grèce, ces courbes stagnent.

4.2. Infractions enregistrées pour 100 000 habitants

Après vous être familiarisé avec le premier graphique, nous allons observer le nombre d'infractions dans ces pays et commenter la potentielle influence envers le premier graphique. Ces valeurs sont toujours pour 100 000 habitants.



Nous pouvons constater que le lien entre le nombre d'infractions enregistrées et la population carcérale n'est pas si évident, cela s'explique sans doute par un nombre assez important de libération dans des pays comme l'Allemagne ou même la France, ou alors que certaines infractions ne font pas forcément l'objet d'emprisonnement. Nous pouvons donc observer que la France possède le nombre d'infractions le plus important parmi les pays concernés (environ 3 500 infractions enregistrées pour 100 000 habitants). On peut aussi voir avec préoccupation que la courbe de la France ne fait qu'augmenter (il y avait environ 2500 infractions pour 100 000 habitants en 2008).

Bien qu'assez élevé en 2008 en Allemagne (environ 4000 infractions enregistrées), la baisse de la courbe se traduit majoritairement par une diminution des vols et des agressions dans le pays au fil du temps et atteint donc 2390 infractions environ en 2018.

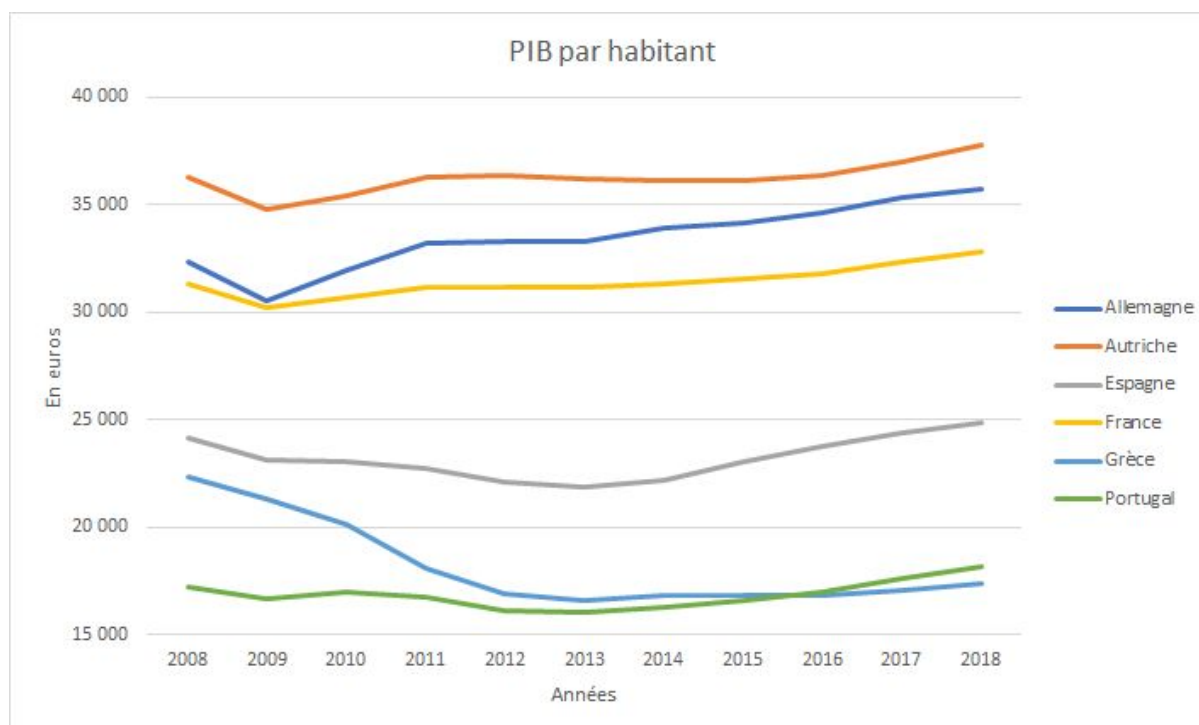
Comme nous pouvons le constater, l'Autriche connaît une situation similaire à l'Allemagne et passe de 3530 infractions enregistrées en 2008 à 2527 en 2018.

La Grèce, l'Espagne et le Portugal se retrouvent quasiment au même point en 2018 avec respectivement 1101, 909 et 1272 infractions et même si les courbes de l'Espagne et de la Grèce stagnent, la courbe du Portugal a diminué puisqu'en 2008 environ 2000 infractions étaient enregistrées.

Après croisement de ces données nous pouvons clairement dire que le nombre d'infractions commises dans un pays n'influence pas majoritairement sa population carcérale (même s'il y contribue), d'autres facteurs sont à prendre en compte comme le taux d'emprisonnement sur ces infractions ou les libérations effectuées.

4.3. PIB par habitant

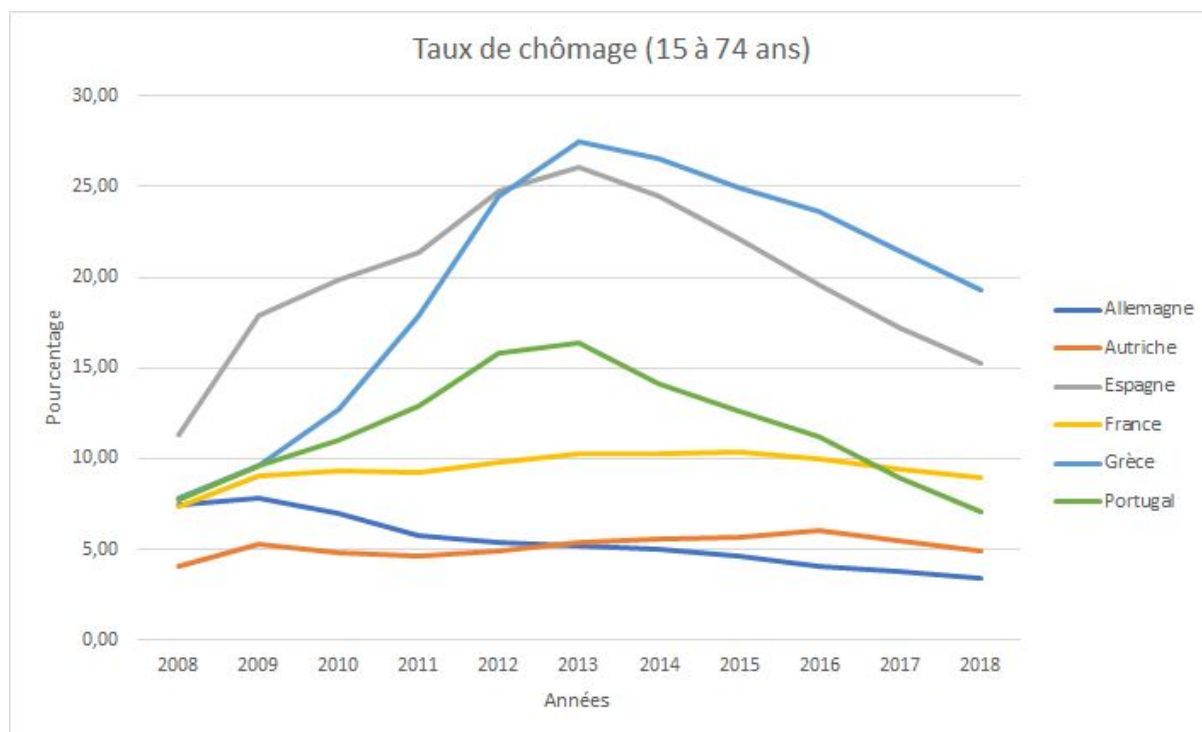
Comme nous l'avons décrit plus haut, le PIB par habitant montre le niveau de vie des différentes populations.



Ensuite, ce graphique doit être vu comme ceci : Nous constatons bien sûr que pour tous les pays, la courbe stagne ou augmente légèrement (on peut quand même noter une baisse de la courbe de la Grèce de 2008 à 2013 à cause des crises successives). Nous avons choisi cet indicateur pour donner une idée générale du niveau de vie des habitants. Ici, les fluctuations des courbes ne sont pas les plus importantes, ce qui nous intéresse c'est que des pays comme l'Allemagne ou l'Autriche possèdent un fort PIB par habitant là et des pays comme la Grèce ou le Portugal en possèdent un moins important. Nous nous servirons de ces "niveaux" pour comparer les données dans les parties suivantes.

4.4. Taux de chômage

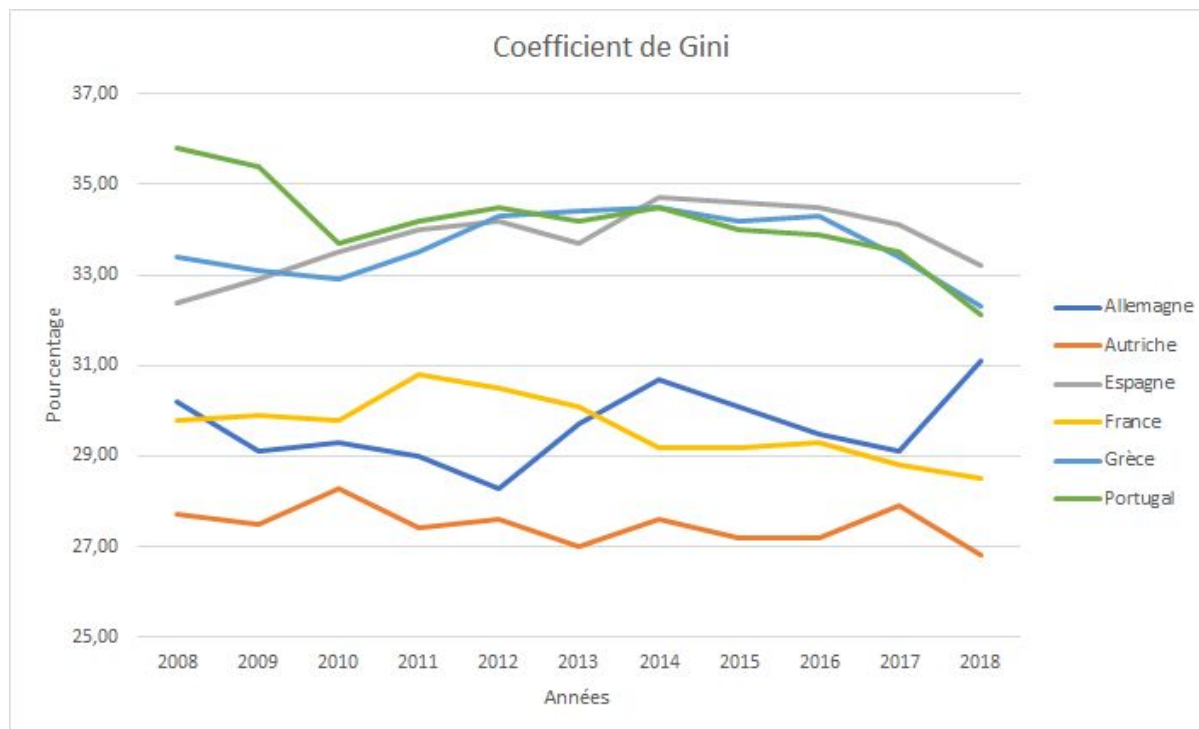
Pour continuer à montrer ce qui peut motiver les infractions dans les pays, nous voulons nous arrêter sur un indicateur social particulier : le taux de chômage des pays. En effet, nous pensons qu'un fort taux de chômage peut encourager le nombre de comportements déviants dans un pays et donc favoriser la quantité d'infractions. La classe d'âge montrée ici est la suivante : 15 à 74 ans.



Si nous revenons dans le cas de l'Allemagne par exemple, nous pouvons voir qu'entre 2005 et 2010 le taux de chômage diminué de façon significative pour le pays puisqu'il passe de 11,2% à 7% en 5 ans. Cette diminution va de pair avec la diminution du nombre d'infractions, mais nous en parlerons plus en détails lors de la partie concernant les croisements des jeux de données.

4.5. Coefficient de Gini

Comme nous l'avons cité précédemment, nous avons sélectionné le coefficient de Gini car il représente les différentes inégalités de richesses entre les pays (ici indiqué en pourcentage).



Tout d'abord, l'indicateur peut se lire de cette façon : un pays ayant un coefficient de Gini de 50 peut signifier que 50% des habitants les plus modestes se partagent 50% des revenus et que 50% des habitants les plus aisés se partagent les autres 50% des revenus de façon égalitaire. De la même manière qu'avec le PIB par habitant, nous nous intéressons plus en détail au niveau des différentes courbes plutôt qu'à leur fluctuation. En effet, à l'exception du Portugal (qui diminue de 5 points en dix ans), aucune évolution majeure n'est à commenter sur la période choisie. Cette façon de procéder est plus pertinente pour la suite de notre développement.

4.6. Présentation des résultats

Cette partie représente la quintessence de notre travail. Il s'agit donc de croiser les données choisies et d'en commenter les résultats.

4.6.1. Allemagne

Dans un premier temps, prenons le cas de l'Allemagne qui possède la population carcérale la plus faible parmi les pays que nous étudions. Leur nombre d'infractions enregistrées est en baisse, bien qu'il reste important (environ 2500 infractions enregistrées pour 100 000 habitants en 2018). Le taux de chômage en Allemagne est passé de 7,5% environ à 3,4% en dix ans, cette baisse et ce taux bas en général explique sans doute une faible population carcérale. Leur PIB par habitant est l'un des plus élevés d'Europe et des pays que l'on étudie. Par contre, les inégalités de revenus sont fortes avec un coefficient de Gini à 31% (0,31) en 2018. Cette tendance à la hausse montre que la population la plus riche du pays se partage le plus de ressources ou encore que les inégalités entre salaires sont assez fortes. La criminalité de l'Allemagne vient donc plus de ses inégalités économiques que de sa situation sociale.

4.6.2. Autriche

Passons maintenant à la situation de l'Autriche. Le nombre de prisonniers est assez similaire à celui de la France. Cependant, la criminalité chute même si elle reste importante. Nous pensons que la chute de la criminalité n'est pas liée qu'au fort niveau de vie des habitants, car il stagne. Le taux de chômage est faible et les inégalités de revenus sont les plus faibles des pays étudiés. Cette diminution des infractions de 2016 à 2018 peut par la légère baisse du coefficient de Gini de 2017 à 2018 et à la légère hausse du PIB par habitant. Cependant, les évolutions sont trop faibles pour pouvoir constater. Nous expliquerons plus en détail la limite de nos indicateurs dans une partie suivante.

4.6.3. Espagne

Croisons maintenant les données avec le cas de l'Espagne. Il est intéressant de noter que ce pays possède la plus grande population carcérale des pays que l'on étudie (à égalité avec le Portugal), même si pour l'Espagne la tendance de la courbe est descendante. Le nombre d'infractions commises est le plus faible des pays que nous étudions et sa tendance penche plutôt sur la régression au fil des années. On remarque donc un recul de la criminalité avec les années avec cependant une forte population carcérale. Pour l'Espagne, il est assez intéressant de comparer les tendances de ce que l'on va corréliser depuis 2008. En effet, l'année 2008 a été synonyme crise financière dans le monde, de ce fait, le PIB par habitant diminue de 25 000 environ à 22 000 de 2008 à 2013 (vers la fin de la *Grande Récession**). Nous remarquons que sur la même période, le taux de chômage augmente de façon significative (+ 15 points environ en 5 ans). De plus, les inégalités entre les riches étaient fortes et sur une tendance ascendante sur la période 2008 à 2013. Malgré ces indicateurs importants, la situation espagnole était plutôt paradoxale durant cette période (on note seulement une hausse des prisonniers de 2008 à 2009), où les indicateurs de criminalité sont à la baisse. Depuis 2013 la situation de l'Espagne s'améliore sur tous les

points. Cependant en 2018, la population carcérale reste forte. Cela peut sans doute s'expliquer par un taux d'incarcération fort, mais aussi par des inégalités entre richesses qui restent fortes (environ 33% pour le coefficient de Gini en 2018). Le taux de chômage, même s'il diminue, reste très élevé en 2018 (environ 15%). Après l'étude de ce pays, à l'aide de nos indicateurs, nous pouvons déduire que l'Espagne est influencée par sa situation économique mais aussi sociale.

4.6.4. La France

Ensuite, prenons le cas de la France, où le nombre de prisonniers stagne à environ 100 pour 100 000 habitants, ce qui est dans la moyenne des pays que nous avons choisi d'étudier. On remarque aussi que le nombre d'infractions en France se multiplie d'année en année et possède le nombre le plus important en 2018 (avec une tendance à la hausse). Le PIB par habitant semble ne pas avoir un impact important sur ces données dans le sens où il est assez élevé. Cependant, le taux de chômage bien que stagnant reste assez élevé pour un pays d'Europe occidentale, nous pouvons donc estimer que cet indicateur est source de tensions en France. En effet, sur dix ans le taux varie autour de 10%. Enfin, en observant le coefficient de Gini de la France, nous pouvons voir que les inégalités restent fortes malgré une diminution au fil du temps. Cependant, si l'on corrèle l'augmentation du PIB par habitant (donc le niveau de vie des habitants) et la diminution des inégalités de revenus, nous pouvons dire que ce qui motive principalement la criminalité en France n'est pas forcément sa situation économique mais plutôt sa situation sociale (où le taux de chômage reste fort).

4.6.5. Grèce

Ensuite, concernant le cas de la Grèce, nous pouvons observer une augmentation du nombre de prisonniers lors de la Grande Récession (2007/2012) avec une baisse du nombre d'infractions enregistrées et une augmentation historique du taux de chômage durant cette période. Durant cette même période, le niveau de vie des habitants s'est réellement dégradé avec une chute du PIB par habitant et une forte présence d'inégalités dans la répartition des richesses qui oscille entre 32 et 34% en 4 ans. Nous remarquons donc que la succession de crises financières dans ce pays n'a pas vraiment impacté sa criminalité. Après cette période, le nombre de prisonniers chute même lourdement qui s'explique sans doute par une hausse des libérations. Depuis la fin des crises, la situation sociale et économique s'améliore légèrement, même si elle reste très préoccupante. L'étude de ce pays nous a montré un faible impact des crises et des tensions internes vis-à-vis de sa criminalité.

4.6.6. Portugal

Enfin, prenons le cas du Portugal. La population carcérale augmente fortement de 2008 à 2012 avec un léger recul du nombre d'infractions enregistrées, on passe respectivement de 100 à 140 prisonniers pour 100 000 habitants en 4 ans et de 2000 à 1600 (environ) infractions enregistrées pour 100 000 habitants en 4 ans. Ce nombre d'incarcération en hausse s'explique par un taux de chômage en hausse pour la même période (on passe de 7.7% à 15.8% en 4 ans). Cela peut aussi s'expliquer avec un niveau

de vie assez faible dans le pays et des inégalités assez fortes. Cependant la corrélation entre le taux de chômage et le nombre d'incarcérés est clairement mis en lien sur cette période (2008 - 2012). Depuis environ 2013, la population carcérale stagne et les infractions enregistrées diminuent légèrement. Le taux de chômage diminue quant à lui drastiquement, et les inégalités entre les richesses diminuent aussi mais restent fortes. Le niveau de vie des habitants augmente légèrement. L'amélioration de ces indicateurs peut expliquer la baisse graduelle du nombre d'incarcérations depuis 2015. L'augmentation du taux de chômage a montré que la criminalité du Portugal a été affectée par sa situation sociale (on remarque une hausse assez forte du nombre d'incarcérations en même temps que l'augmentation du taux de chômage).

4.6.7. Argumentation finale

Pour finir sur le croisement de nos données. Si l'on regarde un pays riche avec un niveau de vie très élevé, et un taux de chômage faible, on peut observer que la criminalité n'y est pas forcément plus faible (on peut même constater le contraire sur la courbe du nombre d'infractions enregistrées), il en va de même pour la population carcérale où certes, l'Allemagne possède la population carcérale la plus faible des pays étudiés mais on peut constater que la France, par exemple, a un nombre de prisonnier supérieur à celui de la Grèce en 2018 . Par l'écriture de ce paragraphe, nous voulons montrer qu'en Europe de l'Ouest, un pays "riche" avec un niveau de vie élevé n'est pas forcément plus rassurant sur sa criminalité qu'un pays moins développé.

En effet, et comme nous avons eu l'occasion de le constater lors du croisement de nos données récoltées. Le taux de chômage, le niveau de vie des habitants ainsi que les inégalités entre les richesses ont un rôle dans l'influence du nombre d'infractions enregistrées ou dans les populations carcérales des pays. Cependant, ce ne sont pas les seuls facteurs déterminants. Il est vrai que nous avons mentionné les crises économiques survenues en 2008 (crise des subprimes) et en 2010 (crise de la dette de la zone Euro) plus tôt dans le développement. Elles ont joué un rôle majeur dans l'augmentation du chômage par exemple. Ces crises ont créé des tensions qui ont pu influencer la criminalité des pays. D'autres paramètres tels que le nombre annuel de libérations par pays n'ont pas été trouvés. Cependant, c'est un facteur à prendre en compte lors de l'étude de l'évolution de la population carcérale.

Enfin, le nombre d'infractions ou la criminalité d'un pays ne se résume pas à son chômage et ses inégalités économiques. Comme nous l'avons mentionné auparavant, les pays ayant le PIB par habitant le plus faible ont un nombre d'infractions faible et ce malgré de fortes inégalités de répartition des richesses. Certains pays ont des tensions sociales plus violentes que d'autres. Le bien-être des populations est aussi à prendre en compte. Chaque pays possède une conjoncture économique et sociale différente et nos données seules montrent seulement les indicateurs et les domaines (économiques, sociaux) qui influencent potentiellement la criminalité et le nombre de prisonniers des différents pays.

4.7. Applications existantes proches

Il est très difficile de trouver des applications existantes proches de notre sujet. Nous n'avons trouvé aucune application Android et ISO. Cependant, il y a un site qui se rapproche fortement de notre sujet : le site <https://www.prison-insider.com> et plus précisément son outil de comparaison. En effet, cet outil permet de comparer des pays et leur situation carcérale. La comparaison permet d'obtenir des données telles que le taux d'homicide des pays ou encore leur population carcérale et son évolution. Prison-insider propose un large nombre d'indicateurs pour un grand nombre de pays, il est donc assez complet. Sa plus grande limite, est le peu d'années disponibles lors des comparaisons. En effet, par exemple pour la France seulement les années 2018 et 2020 sont recensées, c'est aussi le cas pour une majorité des pays d'Europe occidentale. Les années couvertes sont de 2018 à 2020, il n'y a donc aucune visibilité pour les années antérieures à 2018. Cependant, les données Open Data sur lesquelles nous travaillons ont une portée jusqu'à 2018, les données de prison-insider sont donc plus actuelles.

Le site de prison-insider présente les données sous forme de tableaux (qui sont générés à la sélection des pays) et sous forme de catégories (par exemple introduction où on y trouve la population du pays, le taux d'homicide, l'IDH, et le nature du régime gouvernemental). Ce développement a aussi pour but de montrer que prison-insider est un site ergonomique car il est clair, facile à utiliser et facile à lire. Le site est bien organisé et il sait allier données simples et claires avec des descriptions développées.

4.8. Problèmes sur les données

Lors de nos recherches de données, nous avons été confrontés à certaines difficultés pour réunir les différentes informations. Une partie des données que nous aurions voulu étudier, comme le nombre de libérations par pays, furent introuvables. D'autres données en revanche existent sur d'autres sites prônant l'Open Data comme l'UNDP (United Nations Development Programme - Human Development Reports).

En comparant les données présentes sur Eurostat et l'UNDP, nous avons remarqué qu'il y a des contradictions entre les informations des deux sites. Par exemple, si nous prenons la population carcérale de la France pour 100 000 habitants en 2017, nous pouvons voir que la valeur présente sur Eurostat est de 103.25 (donc 104) alors que sur UNDP la valeur est de 106.

Ensuite, nous avons certains manques de données à déplorer. Nous voulions inclure d'autres pays, mais suite à un important manque de données, notamment pour l'Angleterre, Italie, Suisse... Nous nous sommes rabattus sur les pays étudiés dans ce rapport. De plus, lors du calcul des infractions pour la France, les données sur les agressions et les enlèvements sont introuvables en 2017 et 2018, ce qui empêche le calcul du nombre total d'infractions pour ces années. Il y avait également certaines données pour la France où il était indiqué "données estimées". De ce fait, et dans certains cas les données ne sont pas forcément exactes.

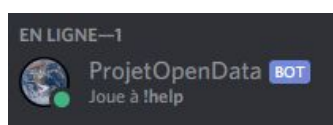
Enfin, nous avons préféré prendre l'intégralité de nos données sur le site de l'Eurostat car après nous être renseignés, nous avons vu que le site Europa étant un site européen, les

données étaient plus précises et exactes contrairement au site de l'UNDP. De plus, les données de UNDP sont téléchargeables au format CSV, ce qui implique une modification de la mise en forme des données afin qu'elles soient plus lisibles et afin de permettre la génération des graphiques. Contrairement au site de de l'UNDP, Europa offre une plus grande variété de format à télécharger, c'est un point positif quant à nos choix de sources de données.

5. Une application

Pour ce projet, nous devons réaliser une application. Après plusieurs idées (Page Web, application...), nous avons opté pour réaliser deux modules différents. Dans un premier temps nous avons développé un logiciel en VBA* qui traite les données présentes sur les feuilles Excel grâce à des macros (ce logiciel fera l'objet d'une étude plus poussée dans la partie suivante).

De plus nous avons eu l'idée d'un bot pour discord qui est un outil, que nous utilisons tous les jours, a été le plus intéressant. Afin de la développer, nous avons utilisé Node.js qui est une plateforme libre en JavaScript permettant l'exécution de code côté serveur, orientée vers les applications réseau.

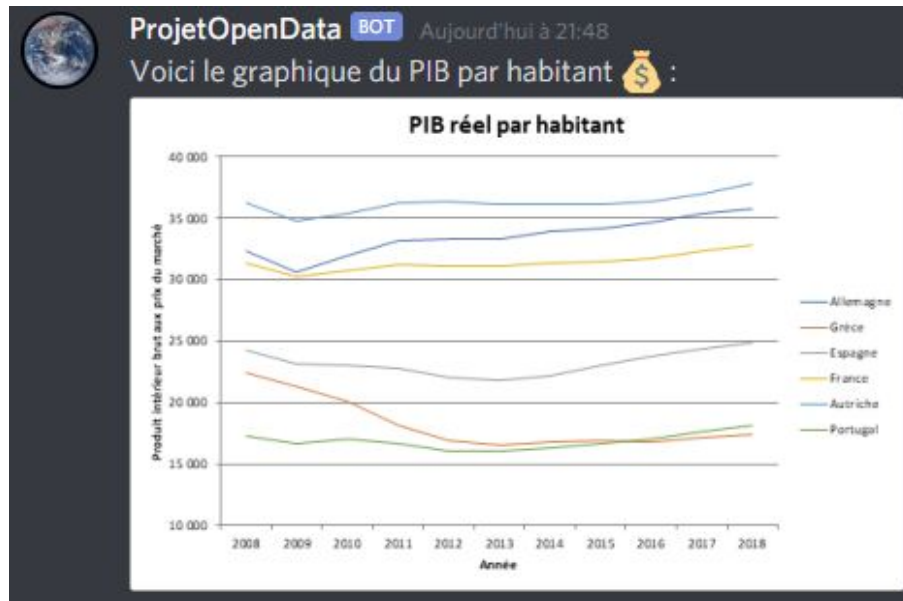


Notre application permet l'affichage des différents graphiques utilisés dans notre projet. Pour cela, plusieurs commandes sont disponibles.

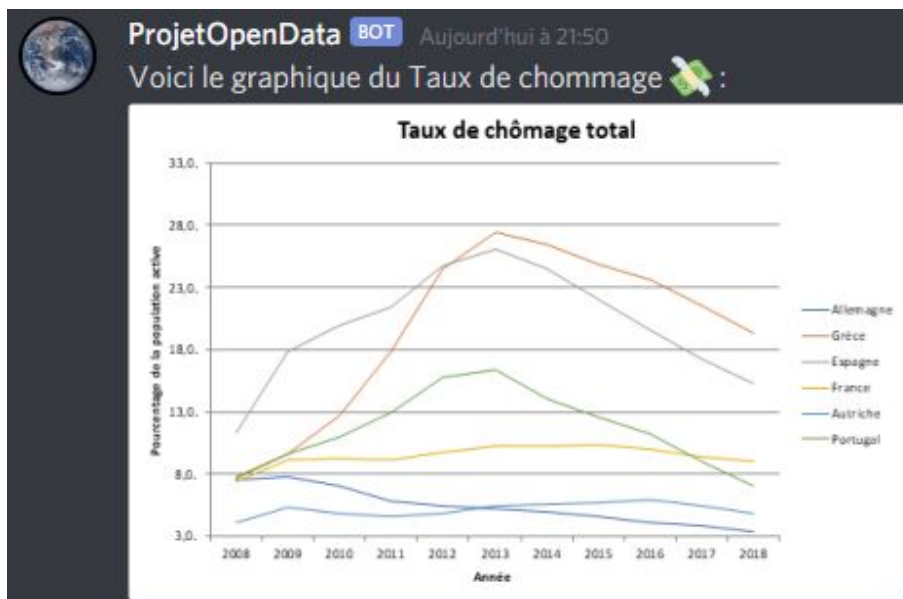
Tout d'abord, il y a la commande principale : "!help", elle permet d'afficher les informations importantes comme : les auteurs, les pays étudiés, ainsi que les différentes commandes utilisables.



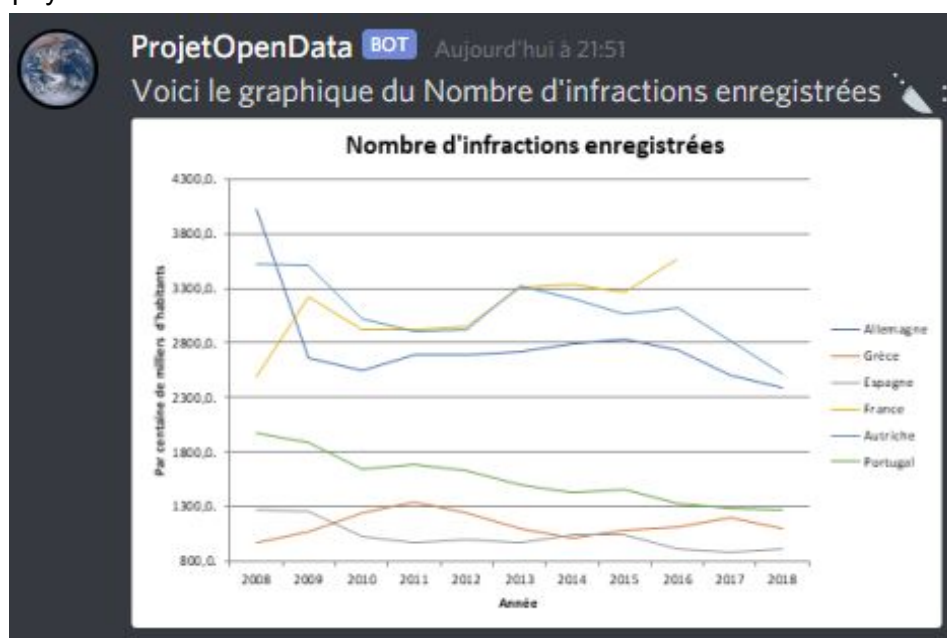
La commande “!pib” va afficher le graphique du PIB par habitant des différents pays.



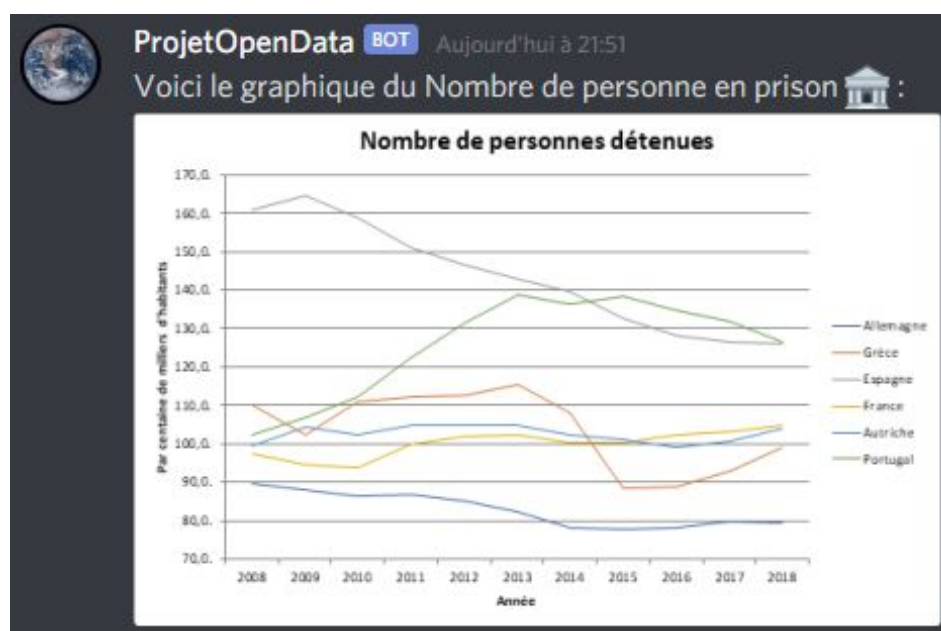
La commande “!txc” va afficher le graphique du taux de chômage dans les différents pays, pour les personnes de 15 à 74 ans.



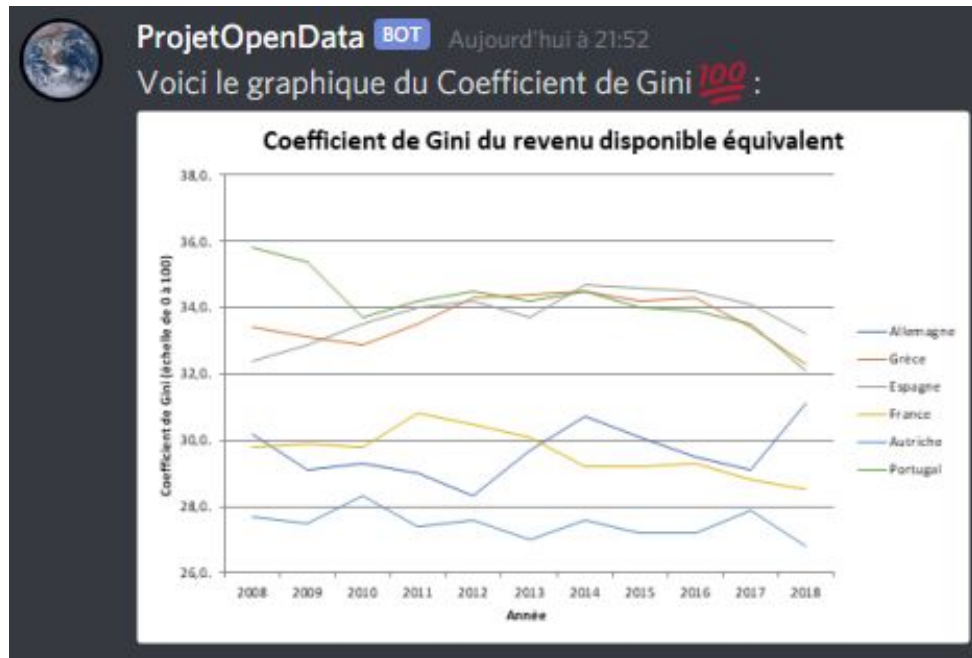
La commande “!nbi” va afficher le graphique du nombre d’infractions enregistrées pour les différents pays.



La commande “!nbp” va afficher le graphique du nombre de prisonniers pour les différents pays.



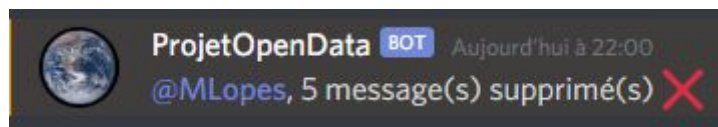
La commande “!cog” va afficher le graphique du coefficient de Gini pour les différents pays.



La commande “!git” va afficher le lien du Github sur lequel se trouve le code source du bot.



Enfin, la commande “!sup X” permet la suppression de X message(s) de la conversion. Elle permet de vider les messages du channel afin de pouvoir supprimer les anciens graphiques de la conversion afin d’en faire apparaître de nouveau par exemple.



Notre application permet donc l’affichage de plusieurs jeux de données sur un serveur discord. Il est donc possible à plusieurs personnes simultanément d’afficher les différents graphiques afin de les avoir sous les yeux et de pouvoir les analyser que ce soit par le chat écrit ou le vocal du serveur. Cette configuration nous permet d’obtenir une interactivité optimale avec les données de notre sujet.

Afin de tester notre bot, il faudra télécharger le code source sur Github. Il sera possible de l’exécuter grâce au lien vers notre discord de test permettant de sauter l’étape 1 du Readme (il faudra tout de même lancer le serveur du bot). Les deux liens se trouvent dans l’annexe section “Webographie”.

6. Présentation des traitements

Concernant le processus du traitement de nos données, comme nous l'avons précédemment énoncé, le site Open Data de l'Europe nous laisse un large choix quant à la sélection de nos données et de leur format avant téléchargement. De ce fait, nous avons choisi de télécharger les données (au format XLS) qui nous intéressent avec un minimum de superflu (sans la mention de certains *flags** à côté de certaines données ou d'autres choses triviales). Nous nous concentrons sur l'essence des données que nous voulons exploiter pour minimiser les modifications post-téléchargement. Ci-dessous une capture d'écran d'un de nos jeux de données après téléchargement :

Données extraites le 21/11/2020 09:39:57 depuis [ESTAT]											
Dataset:	Taux de chômage total [TP\$00203]										
Dernière mise à jour:	07/10/2020 23:00										
Fréquence (relative au temps)	Annuel										
Classe d'âge	De 15 à 74 ans										
Unité de mesure	Pourcentage de la population active										
Sexe	Total										
TIME	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
GEO (Libellés)											
Allemagne (jusqu'en 1990, ancien te	7,5	7,8	7	5,8	5,4	5,2	5	4,6	4,1	3,8	3,4
Grèce	7,8	9,6	12,7	17,9	24,5	27,5	26,5	24,9	23,6	21,5	19,3
Espagne	11,3	17,9	19,9	21,4	24,8	26,1	24,5	22,1	19,6	17,2	15,3
France	7,4	9,1	9,3	9,2	9,8	10,3	10,3	10,4	10	9,4	9
Autriche	4,1	5,3	4,8	4,6	4,9	5,4	5,6	5,7	6	5,5	4,9
Portugal	7,7	9,6	11	12,9	15,8	16,4	14,1	12,6	11,2	9	7,1
Valeur spéciale											
:	Non disponible										

Nous tenons à préciser que cette capture reflète les données après téléchargement et sans modifications de notre part. Pour la suite du traitement, nous avons décidé de supprimer quelques informations qui alourdissent les graphiques. En effet, dans le champ "Dataset : Taux de chômage total [TP\$00203]" nous avons décidé de supprimer ce qui est encadré. De plus, dans le champ Allemagne, nous avons supprimé le contenu des parenthèses qui n'apporte rien à notre sujet. Enfin, nous avons décidé de renommer le champ "Annuel" en "Année" qui apportera un meilleur sens à nos graphiques. Ces modifications sont donc les seuls changements effectués post-téléchargements sur les données. Nous précisons aussi que ces changements ont lieu pour tous les jeux de données exploités.

Il y a une spécification que nous souhaitons apporter au sujet du nombre d'infractions : cette donnée est calculée à partir de 8 types d'infractions différentes. Chaque type d'infraction constitue un jeu de données à part entière. Pour obtenir un montant total d'infraction, nous avons additionné chacun de ces jeux de données pour chaque pays et années :

Cette mise en forme a été réalisée manuellement après avoir réalisé la somme de toutes les infractions.

Données extraites le 21/11/2020 15:16:01 depuis [ESTAT]											
Dataset	Nombre d'infractions enregistrées										
Dernière mise à jour	20/07/2020 23:00										
Fréquence (relative au temps)	Année										
Classification internationale du	Infractions totales										
Unité de mesure	Par centaine de milliers d'habitants										
TIME	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
GEO (Libellés)											
Allemagne	4 027,05	2 664,22	2 544,61	2 696,1	2 691,21	2 727,28	2 794,16	2 841,53	2 739,89	2 506,34	2 390,08
Grèce	973,08	1 072,62	1 244,21	1 337,4	1 249,65	1 096,66	1 016,76	1 090,46	1 111,07	1 200,58	1 100,42
Espagne	1 275,71	1 254,82	1 031,19	977,39	995,17	978,35	1 049,67	1 037,3	916,83	891,34	908,24
France	2 496,76	3 214,88	2 914,61	2 915,53	2 955,87	3 309,07	3 339,51	3 261,52	3 563,07	#N/A	#N/A
Autriche	3 529,96	3 511,47	3 027,98	2 900,58	2 925,86	3 322,91	3 211,91	3 065,2	3 125,39	2 824,58	2 526,61
Portugal	1 979,62	1 885,86	1 649,45	1 690,07	1 637,58	1 504,9	1 432,19	1 464,83	1 333,6	1 292,67	1 271,81
Valeur spéciale											
#N/A	Non disponible										

Avant d'entreprendre la partie processus VBA, il est important de noter que pour exécuter les macros Excel il faut les avoir activées (**fichier-> option -> Centre de gestion de la confidentialité-> Paramètres du Centre de gestion de la confidentialité -> Paramètre des macros-> Activer toutes les macros**).

Nous allons donc passer à la présentation de notre application VBA. Concrètement ici, l'idée a été de générer automatiquement les graphiques en récoltant les bonnes données dans les cellules correspondant à chaque rubrique :


```

Sub create_inf_chart()

' On commence par instancier le graphique avec ses dimensions (il est encore vide)
Set oChartObj = ActiveSheet.ChartObjects.Add(Top:=0, Left:=0, Width:=550, Height:=350)

With oChartObj.Chart
' Ici le diagramme sera sous format de courbes
.ChartType = xlLine
' On active le titre principal du graphique et on lui attribue la valeur
' correspondante (ici dans la feuille du nombre d'infractions
.HasTitle = True
.ChartTitle.Text = Sheets("inf").Cells(2, 2)
' Comme pour au dessus on définit le titre de l'abscisse et de l'ordonnée
.Axes(xlCategory).HasTitle = True
.Axes(xlCategory).AxisTitle.Caption = Sheets("inf").Range("C5")
.Axes(xlValue).HasTitle = True
.Axes(xlValue).AxisTitle.Caption = Sheets("inf").Range("C7")
' On met une valeur minimum sur le graphique pour rendre les courbes plus lisibles
.Axes(xlValue).MinimumScale = Sheets("inf").Range("L1")
' On active ensuite la légende
.HasLegend = True
' Dans la boucle suivante on va attribuer les noms aux pays et les années à l'axe des abscisses
For i = 1 To 6
.SeriesCollection.NewSeries
.SeriesCollection(i).Name = Sheets("inf").Cells(i + 10, 1)
.SeriesCollection(i).XValues = Sheets("inf").Range("B9:L9")
Next i

' On attribue chaque donnée au pays correspondant par ordre des tableaux
' (ici SeriesCollection(1) est l'Allemagne)
' On attribue aussi une couleur personnalisée aux différentes courbes

.SeriesCollection(1).Values = Sheets("inf").Range("B11:L11")
.SeriesCollection(1).Interior.Color = RGB(0, 0, 255)
.SeriesCollection(2).Values = Sheets("inf").Range("B12:L12")
.SeriesCollection(2).Interior.Color = RGB(51, 102, 255)
.SeriesCollection(3).Values = Sheets("inf").Range("B13:L13")
.SeriesCollection(3).Interior.Color = RGB(192, 192, 192)
.SeriesCollection(4).Values = Sheets("inf").Range("B14:L14")
.SeriesCollection(4).Interior.Color = RGB(255, 204, 0)
.SeriesCollection(5).Values = Sheets("inf").Range("B15:L15")
.SeriesCollection(5).Interior.Color = RGB(255, 102, 0)
.SeriesCollection(6).Values = Sheets("inf").Range("B16:L16")
.SeriesCollection(6).Interior.Color = RGB(153, 204, 0)

End With
End Sub

```

Après observation de cette macro (algorithme), nous pouvons observer que la complexité est en $O(n)$ car il y a la présence d'une boucle "for" à l'intérieur (ligne 23).

De plus, nous proposons aussi d'exporter ces graphiques en image sous format PNG. L'export permet au bot de prendre tout son sens en utilisant ces images pour les afficher dans le serveur Discord.

```

Sub export_inf_chart()
' On attribue le nom et le format de l'export avec le chemin de l'enregistrement
graph = ActiveWorkbook.Path & "\\.\\" & "Photos" & "\" & "nbInfractions.png"
ActiveChart.Export Filename:=graph, FilterName:="PNG"
End Sub

```

Voici les boutons qui permettent de contrôler les macros Excel.



Pour conclure cette partie, nous avons eu l'occasion d'utiliser pour la première fois node.js et VBA. Ce fut une expérience intéressante quant au traitement des données (notamment ce qu'offre le VBA quant à la manipulation des données présentes sur Excel).

7. Conclusion

Ce projet fut l'occasion d'obtenir une vision concrète du concept de l'Open Data grâce à la manipulation des données constituant notre travail. Cette partie pratique vient compléter l'approche théorique que nous avons étudiée en cours. L'expérience fut très enrichissante du fait que l'Open Data peut s'avérer être abstrait jusqu'à avoir une approche pratique sur le sujet. Notre principale difficulté fut de trouver des jeux de données pertinents pour constituer un sujet solide qui valait la peine d'être étudié. En effet, certaines données se sont révélées être froides, nous avons aussi constaté l'absence de celles-ci dans certains cas, ce qui a rendu notre tâche du choix du sujet plus difficile encore.

D'un point de vue personnel, l'apport peut être perçu comme un apport de culture générale. Le fait d'étudier le monde carcéral en général (où nous avons trouvé des données assez précises sur la France) nous a vraiment informé sur certains paramètres, comme le nombre important de vols commis en général. Il fut aussi assez ludique de comparer l'influence qu'à la société sur la population carcérale des pays (nous avons utilisé plusieurs types d'indicateurs dans ce but).

Pour finir, le travail sur le nombre de prisonniers des pays fut assez intéressant pour nous bien que compliqué à mettre en œuvre.

ANNEXES

1. Glossaire

Actes illicites liés au domaine de la drogue (d'après europa.eu) : Manipulation, possession, achat, utilisation, trafic, culture ou production illicite de drogues contrôlées ou de précurseurs pour la consommation personnelle et pour la consommation non personnelle. Possession, culture, production, fourniture, transport, importation, exportation, financement, etc. illicites d'opérations liées aux drogues qui ne sont pas uniquement liées à l'usage personnel. Exclusions : Conduite d'un véhicule sous l'influence de substances psychoactives, causant la mort par conduite sous l'influence de drogues ou d'alcool.

Agressions (d'après europa.eu) : Attaque physique envers une personne qui résulte sur des blessures non létales sur la victime. Exclut les homicides, ainsi que les violences sexuelles.

Cambriolages (d'après europa.eu) : L'accès non autorisé à une partie d'un bâtiment, d'un logement ou d'autres locaux, y compris par la force, dans l'intention de voler des biens (effraction). Par "cambriolage", il faut entendre le vol dans une maison, un appartement ou un autre lieu d'habitation, une usine, un magasin ou un bureau, dans un établissement militaire, ou en utilisant de fausses clés.

Copyright : Droit exclusif que détient un auteur ou son représentant d'exploiter une œuvre.

Courbe de tendance polynomiale : Courbe utilisée dans un graphique lorsque les données varient (assez utile pour mesurer des profits ou des pertes dans des jeux de données).

CSV (Comma-Separated Values) : Format de texte ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules.

Enlèvements (d'après europa.eu) : Concerne l'enlèvement d'une personne contre son gré sans but précis (que ce soit pour demander une rançon ou juste pour entraver la personne enlevée). Les prises d'otages sont exclues.

Flags : Ici, par "flags" nous faisons référence à des indicateurs placés autour certaines données Open Data (comme par exemple un flag "e" mentionnant que la donnée en question est une estimation).

Grande Récession : Beaucoup utilisé dans le monde anglophone pour caractériser la crise économique mondiale des années 2007-2012.

Homicide intentionnel (d'après europa.eu): Violence infligée à une personne dans l'intention de causer sa mort ou des blessures graves. Les données sur les homicides intentionnels incluent les meurtres, les crimes d'honneur, les agressions graves entraînant la mort, les décès résultant d'activités terroristes, les meurtres liés à la dot, les féminicides, les infanticides, les homicides volontaires, les exécutions extrajudiciaires, les meurtres causés par un usage excessif de la force par les forces de l'ordre/les agents de l'État. Elles excluent les décès dus à des interventions légales, les homicides justifiables en légitime défense, les tentatives d'homicide intentionnel, les homicides sans élément d'intention, les homicides non intentionnels, les homicides involontaires ou non par négligence, l'aide au suicide ou l'incitation au suicide, l'euthanasie.

HTML (Hypertext Markup Language) : Langage de balisage conçu pour représenter des pages web.

Infraction : Selon le droit pénal, une infraction est une violation d'une loi entraînant des peines fixées par la loi pénale. Dans notre cas, nous avons utilisé huit types d'infractions différentes (celles disponibles en Open Data : homicides, vols, enlèvements, cambriolages, trafic de drogue, violences sexuelles, vols qualifiés, agressions).

JSON (JavaScript Object Notation) : Format de données textuel, il permet de représenter l'information de manière structurée comme le permet XML.

Licence Creative Commons : Constituent un ensemble de licences régissant les conditions de réutilisation et de distribution d'œuvres. Élaborées par l'organisation Creative Commons, elles ont été publiées pour la première fois le 16 décembre 2002.

PC AXIS : Format qui utilise des mots-clés pour structurer l'information.

PDF (Portable Document Format) : Préserve la mise en page d'un document (polices de caractère, images, objets graphiques etc...) telle qu'elle a été définie par son auteur.

Population carcérale : Ici, représente le nombre de détenus dans les prisons des pays que nous étudions.

SDMX (Statistical Data and Metadata eXchange) : Initiative visant à promouvoir la diffusion et les échanges de données et de métadonnées statistiques.

SPSS (SPSS Statistics .sav) : Logiciel développé pour lire des formats comme le TSV, JMP, CSV.

TSV (Tabulation Separated Values) : Format texte ouvert représentant des données tabulaires. Ce format est utilisé pour exporter ou importer des données d'une base de données ou d'un tableur.

URI (Uniform Resource Identifier) : Permet d'identifier les ressources abstraites ou physiques sur internet. Les URI peuvent identifier des sites Web, tout comme des expéditeurs ou des destinataires de courriels.

VBA (Visual Basic for Applications) : Permet de créer des programmes informatiques plus ou moins complexes afin d'automatiser des actions répétitives ou d'effectuer un nombre important d'opérations lors de l'exécution d'une seule commande (macros).

Violences sexuelles (d'après europa.eu) : Acte sexuel non désiré, tentative d'obtenir un acte sexuel, ou contact ou communication avec une attention sexuelle non désirée sans consentement valable ou avec consentement par suite d'intimidation, de force, de fraude, de coercition, de menace, de tromperie, de consommation de drogues ou d'alcool, ou d'abus de pouvoir ou d'une situation de vulnérabilité. Elles excluent les actes d'abus d'une situation de vulnérabilité, de pouvoir ou de confiance, ou l'usage de la force ou de la menace de force, dans le but de tirer un profit monétaire, social ou politique de la prostitution ou des actes sexuels d'une personne, la coercition, les infractions liées à la prostitution, les infractions liées à la pornographie et d'autres actes contraires aux normes sexuelles d'ordre public tels que l'inceste n'équivalant pas à un viol et l'exhibitionnisme, les agressions et les menaces, l'esclavage et l'exploitation n'équivalant pas à des actes préjudiciables de nature sexuelle, la traite des êtres humains à des fins d'exploitation sexuelle, le harcèlement et la traque.

Les chiffres de la violence sexuelle sont la somme des viols et des agressions sexuelles.

Vols : Vols commis sans usage de la force et sans agressions morales. Les vols excluent les cambriolages.

Vols qualifiés : Concerne les vols commis avec agressions physiques ou menaces (vol aggravé).

XLS : Extension de nom de fichier pour tableur au format Excel.

XML (Extensible Markup Language) : Langage de description des données qui utilise un système de balisage générique.

2. Webographie

2.1. Jeux de données

Nombre de personnes en prison :

https://ec.europa.eu/eurostat/web/products-datasets/-/crim_pris_cap

Nombre d'infraction enregistrés :

https://ec.europa.eu/eurostat/fr/web/products-datasets/product?code=crim_off_cat

PIB par habitant :

https://ec.europa.eu/eurostat/web/products-datasets/-/sdg_08_10

Taux de chômage :

<https://ec.europa.eu/eurostat/fr/web/products-datasets/product?code=tps00203>

Coefficient de Gini :

<https://ec.europa.eu/eurostat/web/products-datasets/-/tessi190>

Lien de l'UNDP :

<http://hdr.undp.org/en/data>

2.2. Discord et bot

Lien du discord :

<https://discord.gg/5emBmVaTwN>

Lien du bot:

https://discord.com/api/oauth2/authorize?client_id=775010489427165254&permissions=522336&scope=bot

2.3. Autres liens

Lien du Github :

<https://github.com/mlopesstage/ProjetOpenData-Prison>

Application existante :

<https://www.prison-insider.com/informer/pay>