

*Challenge de création
d'un*
Dataset

avec Scrum

sur le **RSE**

Julie Lallement

Leila Laasry

Ayoub Yacoubi

Mehdi Larbiouene

DM 6 - Quatros Pollos Hermanos

| LES ENTREPRISES ET LE R.S.E |

Dataset provenant d'un jeu de données correspondant aux émissions polluantes des installations industrielles de 2003 à 2020.

| KEY MESSAGE |

La Commission européenne définit la Responsabilité Sociale des Entreprises (RSE) comme "l'intégration volontaire par les entreprises des questions sociales et environnementales dans leurs activités commerciales et leurs relations avec leurs parties prenantes." Afin de contribuer aux enjeux du développement durable, une entreprise qui adopte une stratégie RSE visera à adopter des pratiques plus éthiques et durables dans ses activités.

Le but de ce dataset est d'analyser quel secteur rejette le plus de déchets polluants dangereux, ainsi que dans quelle région. Avec ce jeu de données de plusieurs années, nous pourrions également avoir un suivi de l'évolution de leur production.

Lien du dataset :

<https://drive.google.com/drive/folders/1SSKOnq6R6qJ8DqgAUOJ5qS2dhzzoCizQ?usp=sharing>

Lien du livre blanc Partie 1 :

https://vcs.management-datascience.org/a.alfoce/Quatros_Pollos_Hermanos/src/branch/master/Livre%20blanc/Livre%20blanc%20partie%201

https://vcs.management-datascience.org/a.alfoce/Quatros_Pollos_Hermanos/src/branch/master/Livre%20blanc/Livre%20blanc%20partie%201

Lien du livre blanc Partie 2 :

https://vcs.management-datascience.org/a.alfoce/Quatros_Pollos_Hermanos/src/branch/master/Livre%20blanc

https://docs.google.com/document/d/1Aa4Ce_ovaGxxX6WVa176P8CHPV-ni4EE/edit?usp=sharing&oid=113467457049461362960&rtpof=true&sd=true

Lien du PTT :

<https://docs.google.com/presentation/d/1nEro78crfUQ82DTRQahX8gbVtYHq2Opi/edit?usp=sharing&oid=113467457049461362960&rtpof=true&sd=true>

| KEYWORDS |

RSE, déchets, dangereux, polluants, entreprises

| BACKGROUND |

Nous souhaitons étudier les pratiques RSE au sein des entreprises et leur évolution en fonction des années. Cela demandait un jeu de données très important, mais aussi de savoir quelle partie du RSE nous intéressait. Mais de nombreux sujets font partie du RSE : diversité et discrimination, le bien-être des salariés, la collaboration etc. Nous nous sommes penchés sur l'environnement car c'est l'enjeu principal de notre génération. Ces données nous intéressent car en ayant la possibilité d'étudier l'évolution de la production des déchets polluants et dangereux au fil des années, cela nous permet aussi d'analyser l'évolution de la mentalité des entreprises sur l'environnement.

Elles sont également une ressource importante si l'on veut aider au développement de la réglementation pour réduire les déchets et encourager les mesures de réduction des rejets des polluants, mais aussi diffuser ces informations à quiconque veut se renseigner.

| METHODS |

La Direction générale de la prévention des risques du ministère de l'écologie, du développement durable et de l'énergie recense les principaux rejets et transferts de polluants dans l'eau, l'air et les déchets déclarés par différents établissements, notamment les installations industrielles les plus importantes les stations d'épuration des eaux usées dont la population est supérieure à 100 000 personnes et quelques exploitations d'élevage.

Ces données sont relatives au Registre des Rejets et des Transferts de Polluants (RRTP), un système de collecte et de diffusion de données sur les déversements de matières dangereuses dans l'environnement et les transferts à partir d'établissements industriels et autres.

Lien pour extraire les données :

<https://www.georisques.gouv.fr/donnees/bases-de-donnees/installations-industrielles-rejetant-des-polluants>

Dans ce jeu de données, ce sont des "Parties" (un État ou une organisation d'intégration économique régionale visé à l'article 24 qui a accepté d'être lié par le Protocole sur les registres des rejets et transferts des polluants et pour lequel le Protocole est en vigueur) qui établissent et maintiennent un registre :

(a) Pour les sources ponctuelles, les données sont fournies par établissement ; (b) Pour les sources ponctuelles, les données sont fournies par établissement ; (c) Pour les sources ponctuelles, les données sont fournies par établissement ; (d) Pour les sources ponctuelles, les données sont fournies par établissement.

(a) Dans lequel les données relatives aux sources diffuses peuvent être fournies

(c) Lorsque les données sont fournies par polluant ou par déchet, selon le cas ;

(d) Qui tient compte des divers écosystèmes récepteurs en établissant une distinction entre les rejets dans l'air, le sol et l'eau.

e) Qui fournit des informations sur les transferts ;

f) Est basé sur un système de rapports périodiques obligatoires ;

(g) Se compose de données normalisées et actualisées, ainsi que d'un ensemble de variables normalisées. Comprend, le cas échéant, des données normalisées et actualisées, établit un nombre restreint de seuils de déclaration définis et garantit la confidentialité des données.

(h) Est cohérent et est conçu pour être convivial et accessible au public, y compris sous forme électronique ;

(i) Permet la participation du public à son élaboration et à sa modification ; et

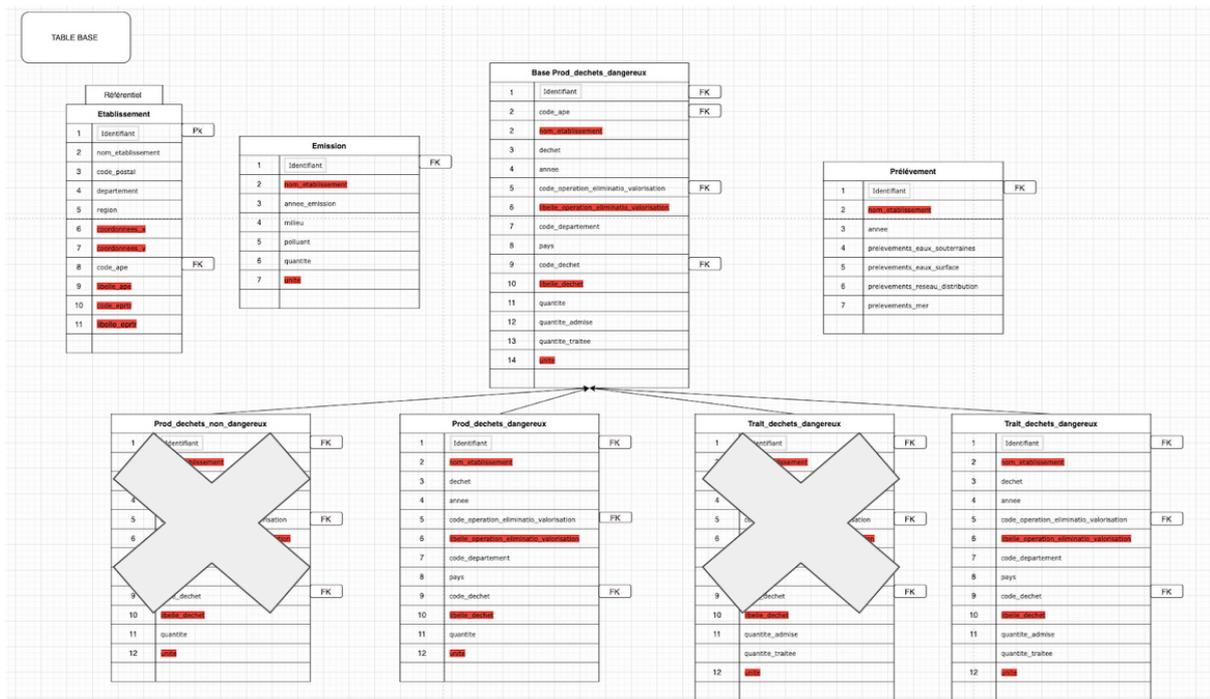
(j) Est constitué de bases de données informatisées, liées et structurées, tenues par l'autorité.

Lien du Protocole : https://treaties.un.org/doc/source/RecentTexts/27_13aF.pdf

| ACCESS TO THE DATA & METADATA DESCRIPTION |

Voici les éléments de base qui composent le registre :

- un rapport annuel sur les rejets et transferts multimédias intégrés (dans l'air, l'eau et le sol) ;
- un rapport par source, couvrant les sources ponctuelles et diffuses, selon le cas ;
- une liste des substances chimiques, familles de substances chimiques ou autres polluants et déchets pertinents qui sont rejetés dans l'environnement ou transférés/traités hors site.
-



Data dictionary :

- Le terme "Établissement" correspond aux entreprises et sociétés sujettes à l'étude.
 - Identifiant : Identifiants correspondants aux installations industrielles
 - Nom_établissement : Etablissements correspondants aux installations industrielles
 - code_postal : Codes postaux correspondants aux installations industrielles
 - departement : Départements correspondants aux installations industrielles
 - région : Régions correspondantes aux installations industrielles
 - code_ape : Code Activité Principale de l'Entreprise permet d'identifier l'activité principale des installations industrielles
- Le terme "Emission" correspond aux niveaux de pollution produite par les entreprises.
 - Identifiant : Identifiants correspondants aux installations industrielles
 - Nom_établissement : Etablissements correspondants aux installations industrielles
 - annee_émission : Corresponds aux emission des installations industrielles lors de l'année N
 - milieu : Correspond à l'environnement contaminé par les déchets polluants (Sol, air ou eau)
 - polluant : Détermine la molécule dangereuse dans les déchets
 - quantité : Correspond à la quantité de déchets polluant produit
 - unité :
- L'expression "Base_Prod_dechets_dangereux" définit le niveau de production de produit dangereux initial des sujets de l'étude.
 - Identifiant : Identifiants correspondants aux installations industrielles
 - code_ape : Code Activité Principale de l'Entreprise permet d'identifier l'activité principale des installations industrielles

- Nom_établissement : Établissements correspondants aux installations industrielles
- dechet :Correspond au type de déchet produit
- annee : Correspond à l'année n
- code_operation_elimination_valorisation : Correspond au code donné à la méthode d'élimination des déchets
- libelle_operation_elimination_valorisation : Correspond à l'intitulé donné à la méthode d'élimination des déchets
- code_departement : Correspond au département des installations industrielles
- pays : Correspond au pays des installations industrielles
- code_dechet : L'identifiant du déchet
- libelle_dechet : Le libellé du déchet
- quantité : Quantité de déchets industriels produits
- quantité_admise : Correspond aux quantité légales de déchets qu'une installation industrielle peut se permettre de produire
- Le terme "Prélèvement" définit le niveau de données prélevées concernant les produits polluants ou non.
 - Identifiant :
 - Nom_établissement : Établissements correspondants aux installations industrielles
 - annee : Correspond à l'année n
 - prelevement_eaux_souterraines : Correspond à la quantité d'eau souterraine prélevée
 - prelevement_eaux_surface : Correspond à la quantité d'eau de surface prélevée
 - prelevement_reseau_distribution : Correspond au réseau de distribution des eaux
 - prelevement_mer : Correspond à la quantité d'eau maritime prélevée
- L'expression "Prod_dechets_dangereux" définit les produit polluant et dangereux pour l'environnement ou les humains
 - Identifiant :
 - Nom_établissement : Établissements correspondants aux installations industrielles
 - dechet :Correspond au type de déchet produit
 - annee : Correspond à l'année n
 - code_operation_elimination_valorisation : Correspond au code donné à la méthode d'élimination des déchets
 - code_departement : Correspond au département des installations industrielles
 - pays : Correspond au pays des installations industrielles
 - code_dechet : L'identifiant du déchet
 - libelle_dechet : Le libellé du déchet
 - quantite : La quantité de dechets dangereux produits par les installations industrielles
- L'expression "Trait_dechets_dangereux" définit la hauteur du traitement des déchets dits dangereux.
 - Identifiant

- Nom_établissement : Établissements correspondants aux installations industrielles
- dechet :Correspond au type de déchet produit
- annee : Correspond à l'année n
- code_operation_elimination_valorisation : Correspond au code donné à la méthode d'élimination des déchets
- libelle_operation_elimination_valorisation : Correspond à l'intitulé donné à la méthode d'élimination des déchets
- code_departement : Correspond au département des installations industrielles
- pays : Correspond au pays des installations industrielles
- code_dechet : L'identifiant du déchet
- libelle_dechet : Le libellé du déchet
- quantite_admise :
- quantite_traite : La quantité de déchets dangereux traitée

| TECHNICAL VALIDATION |

Afin d'améliorer la qualité de notre dataset, nous avons procédé à un nettoyage de la base. Ne disposant pas de machines assez puissantes pour un nombre trop grand de données, le data cleaning nous a permis d'éliminer les données inutiles et de réduire les temps de chargement et de fluidifier les résultats trouvés.

Nous avons commencé par enlever les données qui ne sont pas nécessaires à notre étude, ensuite celles qui représentent des erreurs qui risquent de compromettre le résultat final et finalement les données qui étaient incomplètes.

| REUSE POTENTIAL AND LIMITS |

Ce data sample peut être réutilisé pour faire des études sur le niveau de produits dangereux produits par les entreprises, pour connaître l'évolution de la production de produits dangereux en se focalisant sur un secteur d'activité ou une zone géographique.

| REFERENCES |

Installations industrielles rejetant des polluants. (s. d.). georisques.gouv.fr. Consulté le 13 décembre 2021, à l'adresse

<https://www.georisques.gouv.fr/donnees/bases-de-donnees/installations-industrielles-rejetant-des-polluants>

PROTOCOLE SUR LES REGISTRES DES REJETS ET TRANSFERTS DE POLLUANTS.

(2003, mai). https://treaties.un.org/doc/source/RecentTexts/27_13aF.pdf

ANNEXE : Polluants

N°	Numéro CAS	Polluant	Seuil de rejet (colonne 1)			Seuil de transfert de polluants hors du site (colonne 2) (kg/an)	Seuil de fabrication, de transformation ou d'utilisation (colonne 3) (kg/an)
			Dans l'atmosphère (colonne 1a) (kg/an)	Dans l'eau (colonne 1b) (kg/an)	Dans le sol (colonne 1c) (kg/an)		
1	74-82-8	Méthane (CH ₄)	100 000	-	-	-	*
2	630-08-0	Monoxyde de carbone (CO)	500 000	-	-	-	*
3	124-38-9	Dioxyde de carbone (CO ₂)	100 000 000	-	-	-	*
4		Hydrofluorocarbones (HFC)	100	-	-	-	*
5	10024-97-2	Protoxyde d'azote (N ₂ O)	10 000	-	-	-	*
6	7664-41-7	Ammoniac (NH ₃)	10 000	-	-	-	10 000
7		Composés organiques volatils autres que le méthane (COVNM)	100 000	-	-	-	*
8		Oxydes d'azote (NO _x /NO ₂)	100 000	-	-	-	*
9		Perfluorocarbones	100	-	-	-	*
10	2551-62-4	Hexafluorure de soufre (SF ₆)	50	-	-	-	*
11		Oxydes de soufre (SO _x /SO ₂)	150 000	-	-	-	*
12		Azote total	-	50 000	50 000	10 000	10 000
13		Phosphore total	-	5 000	5 000	10 000	10 000
14		Hydrochlorofluorocarbones (HCFC)	1	-	-	100	10 000
15		Chlorofluorocarbones (CFC)	1	-	-	100	10 000
16		Halons	1	-	-	100	10 000
17	7440-38-2	Arsenic et composés (en As)	20	5	5	50	50
18	7440-43-9	Cadmium et composés (en Cd)	10	5	5	5	5
19	7440-47-3	Chrome et composés (en Cr)	100	50	50	200	10 000
20	7440-50-8	Cuivre et composés (en Cu)	100	50	50	500	10 000
21	7439-97-6	Mercure et composés (en Hg)	10	1	1	5	5
22	7440-02-0	Nickel et composés (en Ni)	50	20	20	500	10 000
23	7439-92-1	Plomb et composés (en Pb)	200	20	20	50	50
24	7440-66-6	Zinc et composés (en Zn)	200	100	100	1 000	10 000
25	15972-60-8	Alachlore	-	1	1	5	10 000
26	309-00-2	Aldrine	1	1	1	1	1
27	1912-24-9	Atrazine	-	1	1	5	10 000

N°	Numéro CAS	Polluant	Seuil de rejet (colonne 1)			Seuil de transfert de polluants hors du site (colonne 2) (kg/an)	Seuil de fabrication, de transformation ou d'utilisation (colonne 3) (kg/an)
			Dans l'atmosphère (colonne 1a) (kg/an)	Dans l'eau (colonne 1b) (kg/an)	Dans le sol (colonne 1c) (kg/an)		
28	57-74-9	Chlordane	1	1	1	1	1
29	143-50-0	Chlordécone	1	1	1	1	1
30	470-90-6	Chlorfenvinphos	-	1	1	5	10 000
31	85535-84-8	Chloroalcane, C ₁₀ -C ₁₃	-	1	1	10	10 000
32	2921-88-2	Chlorpyrifos	-	1	1	5	10 000
33	50-29-3	DDT	1	1	1	1	1
34	107-06-2	1,2-dichloroéthane (EDC)	1 000	10	10	100	10 000
35	75-09-2	Dichlorométhane (DCM)	1 000	10	10	100	10 000
36	60-57-1	Dieldrine	1	1	1	1	1
37	330-54-1	Diuron	-	1	1	5	10 000
38	115-29-7	Endosulphan	-	1	1	5	10 000
39	72-20-8	Endrine	1	1	1	1	1
40		Composés organiques halogénés (en AOX)	-	1 000	1 000	1 000	10 000
41	76-44-8	Heptachlore	1	1	1	1	1
42	118-74-1	Hexachlorobenzène (HCB)	10	1	1	1	5
43	87-68-3	Hexachlorobutadiène (HCBd)	-	1	1	5	10 000
44	608-73-1	1,2,3,4,5,6-hexachlorocyclohexane (HCH)	10	1	1	1	10
45	58-89-9	Lindane	1	1	1	1	1
46	2385-85-5	Mirex	1	1	1	1	1
47		PCDD+PCDF (dioxines + furannes) (en Teq)	0,001	0,001	0,001	0,001	0,001
48	608-93-5	Pentachlorobenzène	1	1	1	5	50
49	87-86-5	Pentachlorophénol (PCP)	10	1	1	5	10 000
50	1336-36-3	Biphényles polychlorés (PCB)	0,1	0,1	0,1	1	50
51	122-34-9	Simazine	-	1	1	5	10 000
52	127-18-4	Tétrachloroéthylène (PER)	2 000	-	-	1 000	10 000
53	56-23-5	Tétrachlorométhane (TCM)	100	-	-	1 000	10 000
54	12002-48-1	Trichlorobenzènes (TCB)	10	-	-	1 000	10 000
55	71-55-6	1,1,1-trichloroéthane	100	-	-	1 000	10 000
56	79-34-5	1,1,2-tétrachloroéthane	50	-	-	1 000	10 000
57	79-01-6	Trichloroéthylène	2 000	-	-	1 000	10 000
58	67-66-3	Trichlorométhane	500	-	-	1 000	10 000
59	8001-35-2	Toxaphène	1	1	1	1	1
60	75-01-4	Chlorure de vinyle	1 000	10	10	100	10 000
61	120-12-7	Anthracène	50	1	1	50	50
62	71-43-2	Benzène	1 000	200 (en BTEX)*	200 (en BTEX)*	2 000 (en BTEX)*	10 000

N ^o	Numéro CAS	Polluant	Seuil de rejet (colonne 1)			Seuil de transfert de polluants hors du site (colonne 2) (kg/an)	Seuil de fabrication, de transformation ou d'utilisation (colonne 3) (kg/an)
			Dans l'atmosphère (colonne 1a) (kg/an)	Dans l'eau (colonne 1b) (kg/an)	Dans le sol (colonne 1c) (kg/an)		
63		Diphényléthers bromés (PBDE)	-	1	1	5	10 000
64		Éthoxylates de nonylphénol (NP/NPE) et substances associées	-	1	1	5	10 000
65	100-41-4	Éthylbenzène	-	200 (en BTEX) ^a	200 (en BTEX) ^a	2 000 (en BTEX) ^a	10 000
66	75-21-8	Oxyde d'éthylène	1 000	10	10	100	10 000
67	34123-59-6	Isoproturon	-	1	1	5	10 000
68	91-20-3	Naphthalène	100	10	10	100	10 000
69		Composés organostanniques (en Sn total)	-	50	50	50	10 000
70	117-81-7	Phtalate de di-(2-éthylhexyl) (DEHP)	10	1	1	100	10 000
71	108-95-2	Phénols (en C total)	-	20	20	200	10 000
72		Hydrocarbures aromatiques polycycliques (HAP) ^b	50	5	5	50	50
73	108-88-3	Toluène	-	200 (en BTEX) ^a	200 (en BTEX) ^a	2 000 (en BTEX) ^a	10 000
74		Tributylétain et composés	-	1	1	5	10 000
75		Triphénylétain et composés	-	1	1	5	10 000
76		Carbone organique total (en C total, ou DCO/3)	-	50 000	-	-	**
77	1582-09-8	Trifluraline	-	1	1	5	10 000
78	1330-20-7	Xylènes	-	200 (en BTEX) ^a	200 (en BTEX) ^a	2 000 (en BTEX) ^a	10 000
79		Chlorures (en Cl total)	-	2 000 000	2 000 000	2 000 000	10 000 ^c
80		Chlore et composés inorganiques (en HCl)	10 000	-	-	-	10 000
81	1332-21-4	Amiante	1	1	1	10	10 000
82		Cyanures (en CN total)	-	50	50	500	10 000
83		Fluorures (en F total)	-	2 000	2 000	10 000	10 000 ^c
84		Fluor et composés inorganiques (en HF)	5 000	-	-	-	10 000
85	74-90-8	Acide cyanhydrique (HCN)	200	-	-	-	10 000
86		Particules (MP ₁₀)	50 000	-	-	-	*

Notes explicatives : Le numéro CAS du polluant renvoie à l'identificateur précis du Chemical Abstracts Service. La colonne 1 indique les seuils visés aux alinéas a i et iv du paragraphe 1 de l'article 7. Si le seuil indiqué dans l'une des sous-colonne (atmosphère, eau ou sol) est dépassé, les Parties qui ont opté pour le système de notification décrit à l'alinéa a du paragraphe 1 de l'article 7, sont tenues de notifier, pour - 32 - l'installation considérée, les rejets ou, s'il s'agit de polluants présents dans des eaux usées destinées à faire l'objet d'une

épuration, les transferts dans le milieu correspondant à cette sous -colonne. La colonne 2 indique les seuils visés à l'alinéa a ii du paragraphe 1 de l'article 7. Si le seuil indiqué dans cette colonne est dépassé pour un polluant donné, les Parties qui ont opté pour le système de notification décrit à l'alinéa a ii du paragraphe 1 de l'article 7 sont tenues de notifier le transfert hors du site de ce polluant, pour l'installation considérée. La colonne 3 indique les seuils visés à l'alinéa b du paragraphe 1 de l'article 7. Si le seuil indiqué dans cette colonne est dépassé pour un polluant donné, les Parties qui ont opté pour le système de notification décrit à l'alinéa b du paragraphe 1 de l'article 7 sont tenues de notifier les rejets et transferts hors du site de ce polluant, pour l'installation considérée. Un tiret (-) indique que le paramètre en question n'entraîne pas une obligation de notification. Un astérisque (*) indique que, pour ce polluant, il convient d'appliquer le seuil de rejet précisé dans la colonne 1 a) de préférence à un seuil de fabrication, de transformation ou d'utilisation. Un double astérisque (**) indique que, pour ce polluant, il convient d'appliquer le seuil de rejet précisé dans la colonne 1 b) de préférence à un seuil de fabrication, de transformation ou d'utilisation. Source : PROTOCOLE SUR LES REGISTRES DES REJETS ET TRANSFERTS DE POLLUANTS