

# ANALYSE DE L'IMPACT ENVIRONNEMENTAL DES ENTREPRISES FRANÇAISES

Thinhinane HAMITOUCHE\* et Harris AOUTCHEME\*\*

\* PSB (Paris School of Business), Paris  
[t\\_hamitouche@stu-psbedu.paris](mailto:t_hamitouche@stu-psbedu.paris)

\*\* PSB (Paris School of Business), Paris  
[h\\_aoutcheme@stu-psbedu.paris](mailto:h_aoutcheme@stu-psbedu.paris)

**Résumé.** Le travail décrit dans cet article a pour but d'analyser l'impact de l'activité des entreprises françaises sur l'environnement, mais également leur engagement en faveur du respect de l'environnement. Nous avons mené une étude analytique en suivant le processus de traitement de la data (collecte, processing, analyse des données et restitution des résultats). Nous avons défini des indicateurs pour construire un tableau de bord qui permettra à chaque entreprise de se positionner. Nos résultats montrent que 23,05% des entreprises de l'échantillon ont un impact « Fort » ou « Très fort » sur l'environnement. Ces dernières sont à 79% des entreprises de 20 à 49 salariés et à 94.8% des PME. Elles sont réparties sur 10 régions en France avec une concentration dans la région Auvergne-Rhône-Alpes. Nous avons également détecté cinq secteurs d'activités qui ne sont pas encore suffisamment engagés comparativement à leur impact. Tous nos résultats sont statistiquement significatifs au seuil de confiance de 95%.

**Mots-clés :** Entreprise française, Environnement, Impact, Engagement.

## 1. Contexte

La question de la protection de l'environnement est devenue une préoccupation majeure, comme en témoignent les nombreuses prises de conscience qui ont abouti à d'importantes évolutions législatives. Le préjudice écologique est désormais inscrit au Code civil. Il est donc essentiel d'intégrer une démarche environnementale dans la stratégie et le management de l'entreprise. Ceci commence par une prise de conscience de l'impact de l'entreprise sur l'environnement.

## 2. Objectifs

Ce travail consiste à analyser l'impact environnemental des entreprises françaises pour les aider à adopter des pratiques de gestion environnementale.

Il est orienté selon les objectifs suivants :

1. Construire un score d'impact et un score d'engagement par secteur d'activité des entreprises en se basant sur l'enquête de l'[INSEE](#) ;
2. Indexer ces scores aux entreprises françaises à partir de la [base Siren](#) ;
3. Organiser et préparer le jeu de données ;
4. Effectuer des analyses statistiques sur les données ;
5. Construire un tableau de bord.

### 3. Etude : processus de traitement de la data

#### 3.1 Collecte des données

Afin de collecter les données, nous nous sommes basés sur six principaux jeux de données (Datasets).

Dataset	Rôle
Siren	Constitue notre Dataset principal. Il regroupe 146575 enregistrements avec des données sur les entreprises françaises, notamment leurs numéros Siren, leurs dénominations, etc
Effectif	Ce Dataset nous renseigne sur les codes des effectifs des entreprises françaises
Localisation	Ce Dataset nous renseigne sur la localisation géographique des entreprises françaises (ville, département, région, etc)
NAF	Ce Dataset nous renseigne sur les codes des secteurs d'activités des entreprises françaises (code NAF à 5 niveaux)
Impact	Ce Dataset représente une enquête sur l'impact des entreprises françaises suivant 8 items en fonction de leurs secteurs d'activités (enquête de l' <a href="#">INSEE</a> )
Engagement	Ce Dataset représente une enquête sur l'engagement des entreprises françaises en faveur de l'environnement suivant 6 items (enquête de l' <a href="#">INSEE</a> )

#### 3.2 Préparation et transformation des données

Nous avons effectué le data cleaning avec Python dans l'environnement Jupyter de la plateforme Management & Datascience. Après importation de tous les Datasets précédemment expliqués, nous avons réalisé un cleaning de chacun d'eux.

##### a) Dataset Siren

La première étape concernant ce Dataset a consisté à supprimer les colonnes que nous avons jugées inutiles. Nous sommes ainsi passés d'un Dataset de 25 colonnes (majoritairement vide) à un Dataset de 5 colonnes contenant uniquement les données pertinentes pour notre étude.

```

#      Column                                     Non-Null Count  Dtype
---  -
0     siren                                       146575 non-null  int64
1     trancheEffectifsUniteLegale                146575 non-null  int64
2     categorieEntreprise                        145210 non-null  object
3     denominationUniteLegale                    146128 non-null  object
4     activitePrincipaleUniteLegale              146575 non-null  object

```

## b) Dataset Effectif

Nous n'avons effectué aucune transformation de ce Dataset car il contient 2 colonnes pertinentes et aucune valeur null.

#	Column	Non-Null Count	Dtype
0	trancheEffectifsEtablissement	14 non-null	int64
1	effectif	14 non-null	object

## c) Dataset Localisation

Sur ce Dataset également, nous avons supprimé les colonnes que nous avons jugées inutiles passant ainsi à 4 colonnes. En examinant le type des colonnes ainsi qu'un échantillon du Dataset, nous avons remarqué que la colonne Siren du Dataset Localisation était de type float64 (avec des 0 dans la partie décimale) tandis que la colonne Siren du Dataset Siren est de type int64. Nous avons donc converti le type de la colonne Siren dans le Dataset Localisation en int64 afin de supprimer cette partie décimale nulle.

#	Column	Non-Null Count	Dtype
0	siren	48648 non-null	int64
1	num_dept	48605 non-null	float64
2	departement	48305 non-null	object
3	region	48305 non-null	object

De plus, ce Dataset contenait une ligne totalement vide qui a été supprimée.

```
df_localisation.loc[df_localisation['siren'].isnull()]
```

	siren	num_dept	departement	region
31058	NaN	NaN	NaN	NaN

## d) Dataset NAF (Niv\_1 et 5\_Niv)

Le Dataset NIV\_1 nous donne le libellé du code NAF qui correspond au niveau 1 (code de niveau 1 et son libellé). Par exemple, les codes A et B correspondent respectivement aux secteurs « industrie agro-alimentaire » et « industries extractives et manufacturières ».

```
df_niv1.head(2)
```

	Code	Libellé
0	A	Industrie agro-alimentaire
1	B	Industries extractives et manufacturières

Le Dataset 5\_NIV nous donne les 5 niveaux du code NAF.

```
df_5_niv.head(10)
```

	NIV5	NIV4	NIV3	NIV2	NIV1
0	01.11Z	1.11	1.1	1	A
1	01.12Z	1.12	1.1	1	A
2	01.13Z	1.13	1.1	1	A
3	01.14Z	1.14	1.1	1	A
4	01.15Z	1.15	1.1	1	A
5	01.16Z	1.16	1.1	1	A
6	01.19Z	1.19	1.1	1	A
7	01.21Z	1.21	1.2	1	A
8	01.22Z	1.22	1.2	1	A
9	01.23Z	1.23	1.2	1	A

Pour obtenir le secteur de l'entreprise en fonction du NIV5 (qui correspond à la colonne « activitePrincipaleUniteLegale » du Dataset Siren), il a fallu joindre les deux Datasets (NIV\_1 et 5\_NIV). Après la jointure, nous avons obtenu le Dataset suivant.

```
df_activite.head(10)
```

	NIV5	NIV4	NIV3	NIV2	NIV1	Code	Libellé
0	01.11Z	1.11	1.1	1	A	A	Industrie agro-alimentaire
1	01.12Z	1.12	1.1	1	A	A	Industrie agro-alimentaire
2	01.13Z	1.13	1.1	1	A	A	Industrie agro-alimentaire
3	01.14Z	1.14	1.1	1	A	A	Industrie agro-alimentaire
4	01.15Z	1.15	1.1	1	A	A	Industrie agro-alimentaire
5	01.16Z	1.16	1.1	1	A	A	Industrie agro-alimentaire
6	01.19Z	1.19	1.1	1	A	A	Industrie agro-alimentaire
7	01.21Z	1.21	1.2	1	A	A	Industrie agro-alimentaire
8	01.22Z	1.22	1.2	1	A	A	Industrie agro-alimentaire
9	01.23Z	1.23	1.2	1	A	A	Industrie agro-alimentaire

## e) Dataset Impact

Nous nous sommes basés sur l'enquête de l'[INSEE](#) pour calculer notre score d'impact environnemental des entreprises. En effet, nous disposons de 12 secteurs d'activités et de 8 items d'impacts. Pour chacun des secteurs, nous avons le nombre d'entreprises ayant un impact de type « Aucun ou très faible », « Faible », « Fort » ou « Très fort » par rapport à chacun des items.

### Comment avons-nous calculé le score d'impact ?

Nous avons retenu pour chaque secteur et chaque item le maximum entre le nombre d'entreprises ayant aucun impact, un faible impact, un fort impact et un très fort impact.

Si ce maximum est égal au nombre d'entreprises ayant un impact de type :

- « Aucun ou très faible » → le score suivant cet item est égal à 1 ;
- « Faible » → le score suivant cet item est égal à 2 ;
- « Fort » → le score suivant cet item est égal à 3 ;
- « Très fort » → le score suivant cet item est égal à 4.

Ainsi, nous avons obtenu un score d'impact pour chacun des 8 items.

Ensuite, nous avons calculé un score d'impact global qui est égal à la somme des 8 scores calculés précédemment. Nous avons remarqué que ce score est compris entre 8 et 12. Nous avons alors construit l'échelle suivante. Si le score d'impact global est égal à :

- 8 → l'impact global est de type « Aucun ou très faible »
- 9 → l'impact global est de type « Faible »
- 10 → l'impact global est de type « Fort »
- 11 ou 12 → l'impact global est de type « Très fort »

#	Column	Non-Null Count	Dtype
0	Secteur	12 non-null	object
1	score_changement_climatique	12 non-null	int64
2	score_la perte_de_la_biodiversité	12 non-null	int64
3	score_pollution_de_air	12 non-null	int64
4	score_pollution_eau	12 non-null	int64
5	score_pollution_sols	12 non-null	int64
6	score_la production_déchets	12 non-null	int64
7	score_substances_toxiques_dangereuses	12 non-null	int64
8	score_nuisances_sonores	12 non-null	int64
9	score_impact_global	12 non-null	int64
10	impact_global	12 non-null	object

Dans la suite de notre étude, nous utilisons les termes score d'impact et impact pour faire référence aux deux colonnes « score\_impact\_global » et « impact\_global » respectivement.

## f) Dataset Engagement

Nous nous sommes basés sur l'enquête de l'[INSEE](#) pour calculer notre score d'engagement des entreprises en faveur de l'environnement. En effet, nous disposons de 12 secteurs d'activités et de 6 items d'engagements. Pour chacun des secteurs, nous avons le pourcentage d'entreprises engagées suivant chacun des items.

### Comment avons-nous calculé le score d'engagement ?

Nous avons retenu pour chaque secteur et chaque item le pourcentage d'entreprises étant engagées. Si ce pourcentage est inférieur ou égal à :

- 25% → le score suivant cet item est égal à 1 ;
- 50% → le score suivant cet item est égal à 2 ;
- 75% → le score suivant cet item est égal à 3 ;
- 100% → le score suivant cet item est égal à 4.

Ainsi, nous avons obtenu un score d'engagement pour chacun des 6 items.

Ensuite, nous avons calculé un score d'engagement global qui est égal à la somme des 6 scores calculés précédemment. Nous avons remarqué que ce score est compris entre 7 et 15. Nous avons alors construit l'échelle suivante. Si le score d'engagement global est égal à :

- 7 → l'engagement global est de type « Aucun ou très faible »
- 8 → l'engagement global est de type « Faible »
- 9 ou 10 → l'engagement global est de type « Fort »
- 11 ou 15 → l'engagement global est de type « Très fort »

#	Column	Non-Null Count	Dtype
0	Secteur	12 non-null	object
1	score_ecolabel	12 non-null	int64
2	score_certificationEnvironementale	12 non-null	int64
3	score_ecoproduit	12 non-null	int64
4	score_ameliorationEfficaciteEnergetique	12 non-null	int64
5	score_gestionRessources	12 non-null	int64
6	score_preservationBiodiversite	12 non-null	int64
7	score_engagement_global	12 non-null	int64
8	engagement_global	12 non-null	object

Dans la suite de notre étude, nous utilisons les termes score d'engagement et engagement pour faire référence aux deux colonnes « score\_engagement\_global » et « engagement\_global » respectivement.

### g) Dataset final

Afin de construire le Dataset final, nous avons fait une jointure de tous les Datasets précédemment présentés, et supprimé les lignes contenant des valeurs null. Nous avons obtenu un Dataset, de 2507 lignes et de 26 colonnes, propre et prêt à être analysé.

```
Data columns (total 26 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   siren                                       2507 non-null   int64
1   denominationUniteLegale                   2507 non-null   object
2   effectif                                   2507 non-null   object
3   categorieEntreprise                       2507 non-null   object
4   Secteur                                    2507 non-null   object
5   num_dept                                   2507 non-null   int64
6   departement                               2507 non-null   object
7   region                                     2507 non-null   object
8   score_changement_climatique               2507 non-null   int64
9   score_la perte_de_la_biodiversité         2507 non-null   int64
10  score_pollution_de_air                    2507 non-null   int64
11  score_pollution_eau                       2507 non-null   int64
12  score_pollution_sols                      2507 non-null   int64
13  score_la production_déchets                2507 non-null   int64
14  score_substances_toxiques_dangereuses     2507 non-null   int64
15  score_nuisances_sonores                    2507 non-null   int64
16  score_impact_global                        2507 non-null   int64
17  impact_global                              2507 non-null   object
18  score_ecolabel                             2507 non-null   int64
19  score_certificationEnvironementale        2507 non-null   int64
20  score_ecoproduit                           2507 non-null   int64
21  score_ameliorationEfficaciteEnergetique   2507 non-null   int64
22  score_gestionRessources                    2507 non-null   int64
23  score_preservationBiodiversite            2507 non-null   int64
24  score_engagement_global                    2507 non-null   int64
25  engagement_global                          2507 non-null   object
```

## 3.2 Analyses des données

Nous avons principalement effectué des tests statistiques qui reposent sur le principe d'inférence, c'est-à-dire le fait de procéder à des généralisations sur les comportements d'une population. En effet, nous avons travaillé sur un échantillon de 2507 entreprises françaises et ces tests nous permettront de généraliser nos résultats à une population plus large.

### a) Test non paramétrique

Nous avons construit des tableaux croisés afin de rassembler dans un tableau unique les distributions de fréquences de deux variables : l'impact de l'entreprise (« Aucun ou très faible », « Faible », « Fort » et « Très fort ») et une autre variable explicative (Secteur d'activité, effectif, statut juridique et région). Ces tableaux sont complétés par des mesures d'association

qui permettent de démontrer la signification statistique d’une association observée entre les variables. Nous avons eu recours au test de Khi\_Deux (Chi<sup>2</sup>).

Nos résultats démontrent une association statistiquement significative, au seuil de confiance de 95% (p <0.05), entre l’impact des entreprises françaises sur l’environnement et leur secteur d’activité, effectif, statut juridique et localisation (voir tableau suivant).

**Test de khi-Deux (Impact & Secteur d'activité)**

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-carré de Pearson	7521.000 <sup>a</sup>	27	.000
Rapport de vraisemblance	5345.355	27	.000
N d'observations valides	2507		

a. 6 cellules (15,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 1,00.

**Test de khi-Deux (Impact & Effectif)**

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-carré de Pearson	134.842 <sup>a</sup>	27	.000
Rapport de vraisemblance	168.574	27	.000
N d'observations valides	2507		

a. 20 cellules (50,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de ,07.

**Test de khi-Deux (Impact & Statut juridique)**

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-carré de Pearson	130.052 <sup>a</sup>	6	.000
Rapport de vraisemblance	168.867	6	.000
N d'observations valides	2507		

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 6,21.

**Test de khi-Deux (Impact & Région)**

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-carré de Pearson	239.826 <sup>a</sup>	36	.000
Rapport de vraisemblance	227.546	36	.000
N d'observations valides	2507		

a. 3 cellules (5,8%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de ,64.

## b) Test paramétrique

Étant donné que nous avons calculé deux scores : score d’impact et score d’engagement, nous comparons les moyennes des deux variables pour voir si la différence est significative.

Pour ce faire, nous avons effectué un test t. Ce dernier démontre une différence de moyennes significative, au seuil de confiance de 95% (p<0.05), entre le score d’engagement et le score d’impact (voir tableau suivant).

**Test T sur échantillon unique**

	t	ddl	Sig. (bilatéral)	Valeur de test = 0 Différence moyenne	Intervalle de confiance de la différence à 95 %	
					Inférieur	Supérieur
Score d'engagement	320,207	2506	,000	8,870	8,82	8,92
Score d'impact	493,253	2506	,000	9,215	9,18	9,25



### 3.3 Tableau de bord

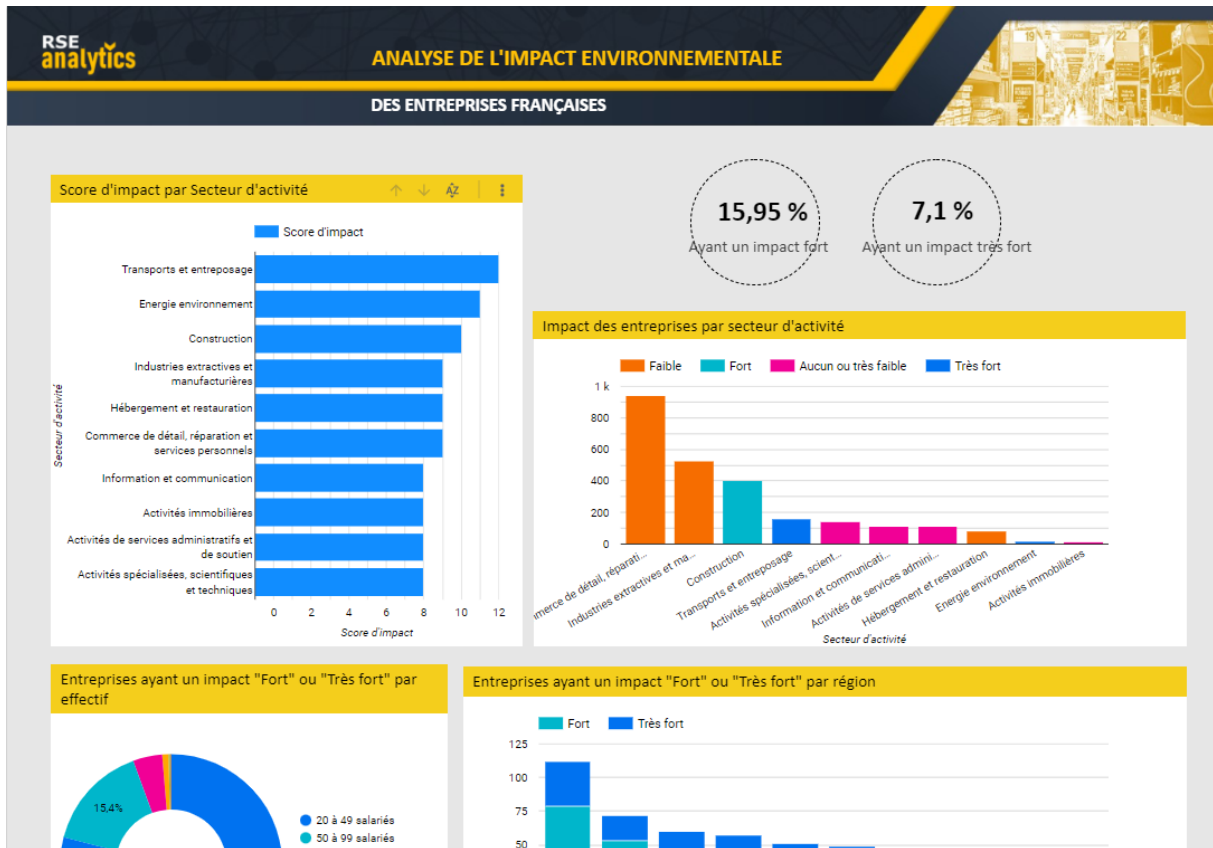
Afin de restituer nos résultats de manière synthétique, nous avons construit un tableau de bord (Dashboard). Ce dernier a pour objectif de sensibiliser les dirigeants au sujet de l'impact environnemental et ainsi, les aider à prendre des décisions relatives à l'adoption des pratiques de gestion environnementale.

Pour ce faire, nous avons défini deux types d'indicateurs :

1. **Indicateurs sur l'impact** : évaluer l'impact de l'activité des entreprises sur l'environnement
2. **Indicateurs sur l'engagement** : évaluer l'engagement des entreprises en faveur de l'environnement

Notre tableau de bord contient les mesures suivantes :

- **Score d'impact par secteur d'activité** : score moyen de l'impact des entreprises de chaque secteur d'activité ;
- **Impact des entreprises par secteur d'activité** : nombre d'entreprises ayant un impact « Aucun ou très faible », « Faible », « Fort » ou encore « Très fort » dans chaque secteur d'activité ;
- **Entreprises ayant un impact « Fort » ou « Très fort » par effectif** : nombre d'entreprises ayant un impact « Fort » ou « Très fort » classées par taille (effectif des salariés) ;
- **Entreprises ayant un impact « Fort » ou « Très fort » par statut** : nombre d'entreprises ayant un impact « Fort » ou « Très fort » classées par statut juridique ;
- **Entreprises ayant un impact « Fort » ou « Très fort » par région** : nombre d'entreprises ayant un impact « Fort » ou « Très fort » classées par localisation, en l'occurrence les régions françaises ;
- **Score d'engagement par secteur d'activité** : score moyen de l'engagement des entreprises de chaque secteur d'activité ;
- **Score d'impact vs score d'engagement des entreprises** : comparaison entre le score moyen de l'impact et le score moyen de l'engagement des entreprises de chaque secteur d'activité.



Le Dashboard est disponible sur [Google Data Studio](#)

#### 4. Discussion

Dans notre étude, en prenant en considération les 8 items d’impacts de leur activité sur l’environnement, les entreprises présentent un score d’impact moyen qui varie selon le secteur d’activité. Les trois scores d’impact les plus élevés, par ordre décroissant, apparaissent dans les secteurs : « transport et entreposage », ensuite « énergie et environnement » et enfin « construction ». Ainsi, 23,05% des entreprises de l’échantillon ont un impact « Fort » ou « Très fort » sur l’environnement.

L’impact environnemental varie également selon la taille de l’entreprise (effectif des salariés), son statut juridique ainsi que sa localisation (ville, département et région). Nos résultats montrent que 79% des entreprises ayant un impact « Fort » ou « Très fort » sont des entreprises de 20 à 49 salariés. Les PME constituent 94.8% des entreprises à impact « Fort » ou « Très fort ». Ces dernières sont réparties sur 10 régions en France avec une concentration dans la région Auvergne-Rhône-Alpes. Ces associations sont statistiquement significatives au seuil de confiance de 95%.

En ce qui concerne le score d’engagement en faveur de l’environnement, nous pouvons classer les trois secteurs les plus engagés par ordre décroissant comme suivant : « énergie et environnement », ensuite « industries extractives et manufacturières » et enfin « hébergement et restauration ».

Nous avons détecté une différence entre le score moyen d'impact et le score moyen d'engagement dans 5 secteurs d'activités, à savoir, « construction », « transport et entreposage », « commerce de détails, réparation et service personnel », « information et communication » et « activités spécialisées, scientifique et techniques ». Nous rappelons que cette différence est statistiquement significative au seuil de confiance de 95%. Elle est pertinente pour les deux secteurs « construction » et « transport et entreposage » qui sont classés parmi les secteurs les plus impactants.

## **5. Conclusion**

Cette étude nous a permis d'estimer l'impact environnemental des entreprises françaises. Grâce au tableau de bord mis en place, nous avons pu mettre en évidence l'impact en fonction du secteur d'activité des entreprises, de leur taille, de leur statut juridique et de leur localisation, ainsi que l'engagement pris en faveur du respect de l'environnement. Cet outil a pour objectif de sensibiliser les dirigeants au sujet de l'impact environnemental et ainsi, les aider à prendre des décisions relatives à l'adoption des pratiques de gestion environnementale. En effet, cinq secteurs ne sont pas encore assez engagés comparativement à leur impact, notamment les secteurs « construction » et « transport et entreposage ».

## **Annexe**

Ressources disponibles sur ce [GitHub](#)