

## French Industry

**Cursus concerné : Data Analyst**

**Niveau de difficulté : 06/10**

### Description du projet :

*Contexte : L'INSEE est l'institut officiel français qui collecte des données de tous types sur le territoire français. Elles peuvent être démographiques (Naissances, Décès, Densité de la population...), économiques (Salaires, Entreprises par activité / taille...) et plus encore.*

*Ces données peuvent être d'une grande aide pour observer et mesurer les inégalités au sein de la population française.*

*Objectif :*

*Comparer les inégalités en France :*

- *Entreprises en fonction de leur localisation, de leur taille.*
- *Population en fonction du salaire et de la localisation.*
- *Focus sur une grande ville*

### Ressources à consulter :

- **Données :**
  - [https://assets-datascientest.s3.eu-west-1.amazonaws.com/notebooks/power\\_bi/power\\_bi\\_datasets\\_projet/french\\_industry.zip](https://assets-datascientest.s3.eu-west-1.amazonaws.com/notebooks/power_bi/power_bi_datasets_projet/french_industry.zip)

### Description des tables

- **base\_etablissement\_par\_tranche\_effectif.csv** : Informations sur le nombre d'entreprises dans chaque ville française classées par taille.
  - CODGEO : ID géographique de la ville
  - LIBGEO : nom de la ville
  - REG : numéro de région
  - DEP : numéro de département
  - E14TST : nombre total d'entreprises dans la ville
  - E14TS0ND : nombre d'entreprises de taille inconnue ou nulle dans la ville
  - E14TS1 : nombre d'entreprises de 1 à 5 employés dans la ville
  - E14TS6 : nombre d'entreprises de 6 à 9 employés dans la ville

[DataScientest.com](https://www.datascientest.com)

Agrément organisme de formation 11755665975

09 80 80 79 49

2 place de Barcelone, 75016 Paris

- E14TS10 : nombre d'entreprises de 10 à 19 employés dans la ville
- E14TS20 : nombre d'entreprises de 20 à 49 employés dans la ville
- E14TS50 : nombre d'entreprises de 50 à 99 employés dans la ville
- E14TS100 : nombre d'entreprises de 100 à 199 employés dans la ville
- E14TS200 : nombre d'entreprises de 200 à 499 employés dans la ville
- E14TS500 : nombre d'entreprises de plus de 500 employés dans la ville
- **name\_geographic\_information.csv** : Données géographiques sur les villes françaises (principalement la latitude et la longitude, mais aussi les codes et les noms des régions/départements).
- **net\_salary\_per\_town\_categories.csv** : Salaires par villes française par catégories d'emploi, âge et sexe
  - CODGEO : ID géographique de la ville
  - LIBGEO : nom de la ville
  - SNHM14 : salaire net moyen
  - SNHMC14 : salaire net moyen par heure pour les cadres
  - SNHMP14 : salaire net moyen par heure pour un cadre moyen
  - SNHME14 : salaire net moyen par heure pour l'employé
  - SNHMO14 : salaire net moyen par heure pour le travailleur
  - SNHMF14 : salaire net moyen pour les femmes
  - SNHMF14 : salaire net moyen par heure pour les cadres féminins
  - SNHMF14 : salaire net moyen par heure pour les cadres moyens féminins
  - SNHMF14 : salaire net moyen par heure pour une employée
  - SNHMF14 : salaire net moyen par heure pour une travailleuse
  - SNHMH14 : salaire net moyen pour un homme
  - SNHMH14 : salaire net moyen par heure pour un cadre masculin
  - SNHMH14 : salaire net moyen par heure pour les cadres moyens masculins
  - SNHMH14 : salaire net moyen par heure pour un employé masculin
  - SNHMH14 : salaire net moyen par heure pour un travailleur masculin
  - SNHM1814 : salaire net moyen par heure pour les 18-25 ans
  - SNHM2614 : salaire net moyen par heure pour les 26-50 ans
  - SNHM5014 : salaire net moyen par heure pour les >50 ans
  - SNHMF1814 : salaire net moyen par heure pour les femmes âgées de 18 à 25 ans
  - SNHMF2614 : salaire net moyen par heure pour les femmes âgées de 26 à 50 ans
  - SNHMF5014 : salaire net moyen par heure pour les femmes de plus de 50 ans
  - SNHMH1814 : salaire net moyen par heure pour les hommes âgés de 18 à 25 ans



# DataScientest

- SNHMH2614 : salaire net moyen par heure pour les hommes âgés de 26 à 50 ans
- SNHMH5014 : salaire net moyen par heure pour les hommes de plus de 50 ans
- **population.csv** : Informations démographiques par ville, âge, sexe et mode de vie

NIVGEO : geographic level (arrondissement, communes...)

CODGEO : unique code for the town

LIBGEO : name of the town

MOCO : mode de cohabitation :

- 11 = enfants vivant avec deux parents
  - 12 = enfants vivant avec un seul parent
  - 21 = adultes vivant en couple sans enfant
  - 22 = adultes vivant en couple avec enfants
  - 23 = adultes vivant seuls avec enfants
  - 31 = personnes étrangères à la famille vivant au foyer
  - 32 = personnes vivant seules
- AGE80\_17 : catégorie d'âge (tranche de 5 ans) | ex : 0 -> personnes âgées de 0 à 4 ans

SEXE : sexe, 1 pour homme | 2 pour femme

NB : Nombre de personnes dans la catégorie

## Conditions de validation du projet :

- un **rapport** d'exploration, de data visualisation et de pre-processing des données ;
- un **rapport** de modélisation ;
- un **rapport** final et le **code** associé.

[DataScientest.com](https://DataScientest.com)

Agrément organisme de formation 11755665975

09 80 80 79 49

2 place de Barcelone, 75016 Paris