

HACKATON 2025

INTRODUCTION

PROJET CHOISI



Plan France Relance

Quels impacts des aides distribués sur
l'économie et la transition écologique ?



**DANS QUELLES MESURES LES AIDES
FINANCIÈRES DU PLAN DE RELANCE ONT-ELLES
CONTRIBUÉ À LA RÉSILIENCE DES ENTREPRISES
FRANÇAISES FACE À LA CRISE DU COVID-19 ?**

PRÉSENTATION DES DONNÉES

- **Siren**
- **dateCréation**
- **trancheEffectifs**
- **catégorieEntreprise**
- **etatAdministratif**
- **nom**
- **activitéPrincipale**
- **catégorieJuridique**
- **nomenclatureActivitéPrincipale**
- **année**

PRÉSENTATION DES DONNÉES



ANALYSE DES DONNÉES

```
df.shape
```

```
(98210861, 11)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 98210861 entries, 0 to 98210860
Data columns (total 11 columns):
 #   Column           Dtype  
 --- 
 0   siren            object  
 1   dateCreationUniteLegale    object  
 2   trancheEffectifsUniteLegale    object  
 3   categorieEntreprise        object  
 4   dateDebut             object  
 5   etatAdministratifUniteLegale    object  
 6   nomUniteLegale          object  
 7   activitePrincipaleUniteLegale    object  
 8   categorieJuridiqueUniteLegale    float64 
 9   nomenclatureActivitePrincipaleUniteLegale    object  
 10  nicSiegeUniteLegale        float64 
dtypes: float64(2), object(9)
```

ANALYSE DES DONNÉES

```
# Visualisation des nan
for i in df.columns:
    print(i, f" : {((df[i].isna().sum()) / len(df)) *100} %")

siren : 0.0 %
dateCreationUniteLegale : 71.90950907150687 %
trancheEffectifsUniteLegale : 70.71582744804569 %
categorieEntreprise : 89.26170090291745 %
dateDebut : 0.8302890247546043 %
etatAdministratifUniteLegale : 2.2478929290722744 %
nomUniteLegale : 83.46302350409086 %
activitePrincipaleUniteLegale : 70.73774559414564 %
categorieJuridiqueUniteLegale : 70.71582744804569 %
nomenclatureActivitePrincipaleUniteLegale : 70.73774559414564 %
nicSiegeUniteLegale : 70.71582744804569 %
```

```
# Suppression des nan non problématique
df = df[df["etatAdministratifUniteLegale"].isna() == False]
df = df[df["dateDebut"].isna() == False]
```

df.shape

(95187752, 11)

- **PROPAGATION DU DERNIER STATUT SUR LES AUTRES VALEURS/ANNÉES**

CONCLUSION

- **TRAVAILLER SUR LES
DONNÉES GÉOGRAPHIQUES**
- **RÉALISER UN MODÈLE DE
MACHINE LEARNING**
- **TROUVER DES CSV QUI
AVAIENT DES DONNÉES
CROISÉES**